Boston Studies in the Philosophy of Science 292

Léna Soler
Emiliano Trizio
Thomas Nickles
William C. Wimsatt *Editors*

Characterizing the Robustness of Science

After the Practice Turn in Philosophy of Science

Springer

CHARACTERIZING THE ROBUSTNESS OF SCIENCE

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 292

For further volumes:
http://www.springer.com/series/5710

CHARACTERIZING THE ROBUSTNESS OF SCIENCE

After the Practice Turn in Philosophy of Science

Edited by

LÉNA SOLER
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, Nancy, France

EMILIANO TRIZIO
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, Nancy, France

THOMAS NICKLES
University of Nevada, Reno, NV, USA

and

WILLIAM C. WIMSATT
University of Chicago, Chicago, IL, USA

Springer

Editors
Léna Soler
Archives H. Poincaré, LHPS
91, avenue de la Libération, 54000
Nancy
France
l_soler@club-internet.fr

Emiliano Trizio
Archives H. Poincaré, LHPS
91, avenue de la Libération, 54000
Nancy
France
emilianotrizio@hotmail.com

Thomas Nickles
Department of Philosophy
University of Nevada-Reno
Reno, NV
USA
nickles@unr.edu

William C. Wimsatt
Department of Philosophy
University of Chicago
Chicago, IL
USA
wwim@uchicago.edu

ISSN 0068-0346
ISBN 978-94-007-2758-8 e-ISBN 978-94-007-2759-5
DOI 10.1007/978-94-007-2759-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011945063

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To the memory of Cathy Dufour, our very dear friend and colleague, a bright, dynamic, and generous person, who prematurely and suddenly passed away in March 2011. She was a pivotal contributor to the PratiScienS research devoted to robustness.

Contents

Contributors

Catherine Allamel-Raffin IRIST, University of Strasbourg, Strasbourg, France, allamelraffin@unistra.fr

Mieke Boon Department of Philosophy, University of Twente, Enschede, The Netherlands, M.Boon@utwente.nl

Catherine Dufour† P2M Department, Institut Jean Lamour, UMR 7198, Université Henri Poincaré, Nancy, France

Jean-Luc Gangloff IRIST, University of Strasbourg, Strasbourg, France, gangloff@unistra.fr

Ralf Krömer University of Siegen, Siegen, Germany, kroemer@mathematik.uni-siegen.de

Hubertus Nederbragt Descartes Centre for the History and Philosophy of the Sciences and the Humanities, Utrecht University, Utrecht, The Netherlands, h.nederbragt@planet.nl

Thomas Nickles Department of Philosophy, University of Nevada, Reno, NV, USA, nickles@unr.edu

Andrew Pickering Department of Sociology and Philosophy, University of Exeter, Exeter EX4 4RJ, UK; Department of Sociology, Kyung Hee University, Seoul, Korea, a.r.pickering@exeter.ac.uk

Léna Soler Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France, l_soler@club-internet.fr

Jacob Stegenga University of Toronto, Toronto, ON, Canada, jacob.stegenga@utoronto.ca

Emiliano Trizio Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France; Archives Husserl, Paris, France; Department of Philosophy, Seattle University, Seattle, WA, USA, emilianotrizio@hotmail.com

Frédéric Wieber Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France, frederic.wieber@wanadoo.fr

William C. Wimsatt Department of Philosophy and Conceptual and Historical Studies of Science, The University of Chicago, Chicago, IL, USA; Center for Philosophy of Science, University of Minnesota, St. Paul, MN, USA, wwim@uchicago.edu

Chapter 1
Introduction: The Solidity of Scientific Achievements: Structure of the Problem, Difficulties, Philosophical Implications

Léna Soler

The disciplines whose scientific status is not in question, such as physics, are characterized by what is commonly described as the 'reliability' and 'successfulness' of their theoretical, experimental or technical accomplishments. Today, philosophers of science often talk of "robustness".

At first sight, robustness seems to be an intuitively clear concept. Yet, it is far from easy to give a precise account of just what is implied by this notion. Major fluctuations and vagueness in the use of robustness terminology only accentuates this substantial difficulty. As stressed by one of the contributors to this book, Thomas Nickles: "In recent years the term 'robustness', its cognates and neighbors (solidity, persistence, hardiness, reliability, resilience, viability, flexibility, healthiness, etc.) have been applied to just about everything. In fact, 'robust' has become a buzzword in popular culture that can be applied to anything that exhibits strength of some sort." Indeed, the word is commonly applied, in addition to achievements in the scientific field (scientific propositions, scientific theories, experimental devices, experimental results and the like), also to complex technological systems beyond the scientific field (aircraft, electrical grids, nuclear plants, the Internet and the like), as well as to biological organisms and to social networks. So there is a need for both substantive investigation and terminological clarification.

At the most general level, central questions await further exploration: What exactly lies at the basis of the robustness of the various sciences? How is robustness historically generated and improved upon? What does it mean that a scientific result or a developmental stage of science is 'more robust' than another? Behind such questions lurk crucial epistemological issues: indeed, nothing less than the very nature of science and its specificity with respect to other human practices, the nature of rationality and of scientific progress, and (last but not least) science's claim to be a truth-conducive activity.

L. Soler (✉)
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie,
UMR 7117 CNRS, Nancy, France
e-mail: l_soler@club-internet.fr

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_1,

In relation to these questions, William Wimsatt has been a pioneer, and his writings constitute a fundamental reference point. The first chapter of the present volume reprints his seminal 1981 paper, "Robustness, Reliability, and Overdetermination", which has been the point of departure for work on robustness since that time.[1] In a new chapter that appears in this volume for the first time, Wimsatt provides a personal account of the peculiar situation in evolutionary biology that he experienced as a young professor in the 1970s, which sparked his lifelong concern with robustness and which had already motivated authors such as Richard Levins and Donald Campbell to introduce notions related to robustness.

In the 1981 article reprinted in the present book, Wimsatt developed a systematic general analysis inspired by these attempts and similar ones, that he called "robustness analysis". Through this article, he introduced into philosophy of science the informal robustness vocabulary already widely appearing in scientists' own talk and proceeded to develop a more specific and technical one that, while preserving the common association with the ideas of reliability and successfulness, allows a more precise characterization. Robustness is defined as the use of "multiple means of determinations" to "triangulate" the existence and the properties of a phenomenon, of an object or of a result. The fundamental idea is that any object (a perceptual object, a physical phenomenon, an experimental result, etc.) that is sufficiently invariant under several independent derivations (in a wide sense of the term 'derivation', including means of identification, sensorial modalities, measurements processes, tests, models, levels of description, etc.) owes its strength (i.e. its robustness) to this situation.

Historically, the question of the reliability of science was first formulated, within philosophy of science, as a problem concerning the relations between statements. The so-called "practice turn" of Science Studies that began in the 1980s has produced a shift in focus. The practice turn has led to an enlarged characterization of science that includes several aspects previously ignored or underplayed on the grounds of their alleged epistemological irrelevance (tacit knowledge, local norms and standards, instrumental resources and the like). Since the practice turn showed that these aspects are often not anecdotal, the investigation of the issue of robustness, and the resulting characterization, must take them into account. This is what the reference to "the practice turn" in the sub-title of this book is intended to mean.

To open this volume on robustness in science I would like, starting from William Wimsatt's writings on robustness, to propose a general analysis of what will be called 'the problem of the solidity of scientific achievements'. (The reason why the term 'solidity' replaces the more common term 'robustness' will appear in the next section.) First the problem and its structure, as well as terminological clarifications and suggestions, are introduced. Then several difficulties that any robustness analysis inspired by Wimsatt's framework will have to face, are identified and characterized. The philosophical issues related to the solidity problem are also clarified.

[1] Thanks to William Wimsatt for offering to re-print this fundamental chapter in the present book. It is all the more important because this chapter is a pivotal source of inspiration for most of the contributions of this book, and is very often quoted or exploited. So the reader will have the original chapter at hand.

Along the way, I shall indicate which chapter of this book contributes to the investigation of these difficulties and philosophical issues. Finally, a more systematic and sequential overview of the different chapters of this book will be provided.

1.1 Robustness. . . That Is to Say?

Let me begin with some terminological remarks, which, clearly, are not just a matter of words.

As already suggested above, the term 'robustness', central to the volume title, is, today, very often employed within philosophy of science in an intuitive, non-technical and flexible sense that, globally, acts as a synonym of 'reliable', 'stable', 'effective', 'well established', 'credible', 'trustworthy', or even 'true'. Correlatively, the more precise and technical sense developed by Wimsatt refers to the idea of the invariance of a result under multiple independent determinations. However, as Wimsatt himself explicitly admits, the scheme of invariance under independent multi-determinations does not exhaust the ways in which an element of scientific practices can acquire the status of a 'robust' ingredient in the broad sense of the term.

In order to avoid confusion between the intuitive sense and Wimsatt's technical sense, and to make room for other possible schemes of constitution of reliability, it would be desirable to have at hand an appropriate specific standard terminology shared by the members of the Science Studies community. To take a step in this direction, I suggest the following terminological options.

I will restrict the term 'robustness' to the specific sense introduced by Wimsatt, that is:

> X is robust $=$ X remains invariant under a multiplicity of (at least partially) independent derivations.

In order to capture the broader, more general and largely indeterminate sense found in common use, I will employ the term 'solidity'.

On the basis of these decisions, robustness appears as a particular case of solidity[2]:

> X is robust $=$ X is solid, *for* X remains invariant under partially independent multiple determinations.

[2] Several authors of this book have adopted my terminological proposal, but of course, alternative terminological options are possible, as soon as they are clearly specified. For example Mieke Boon, in her paper, chooses to retain the expression "robustness notions" to encompass all the various uses connected to the intuitive idea of robustness. Then she decomposes this umbrella expression into different but related features (such as reality, reproducibility, stability etc.) that can be attributed to specified kinds of things (independent reality, physical occurrences, observable and theoretical objects, etc.), features that may differ in status (metaphysical, ontological, epistemological, etc.). In such a framework robustness is a generic notion, and we then have different species of robustness (robustness in the sense of reliability, robustness in the sense of multiple determinations, etc.).

However, other schemes could be involved in the process of constituting solidity. For instance, Wimsatt has also insisted on another scheme, different from that of robustness, that he has called "generative entrenchment"[3] (hereafter, GE):

> *X* is GEed = *X* is solid, *for X* is involved in an essential way (*X* plays a quasi-foundational role) in the generation of a huge number of ingredients constituting scientific practices.

I propose the term of 'solidity' for two reasons. First, it is not already associated, as far as I know, with any specific, technical sense in Science Studies. Second, its meaning in ordinary usages is very broad: the term is currently applied to very heterogeneous kinds of things, either material or intellectual. This second characteristics is, as we will see (Section 1.7.3), a desirable feature if we want to be in a position to think of the robustness of science in relation to scientific *practices*. Indeed, many heterogeneous kinds of elements are involved in these practices and can be implicated, in one way or another, in the generation of something that we are intuitively inclined to call 'robust' or 'solid'. On the basis of the terminology just proposed, the title of this volume should be changed to: *Characterizing the* Solidity *of Science After the Practice Turn in Philosophy of Science*. The editors nevertheless choose to retain the initial title, since 'robustness' is a more familiar descriptive term for people who work in Science Studies.

No doubt there exist other schemes of the acquisition of solidity besides robustness and generative entrenchment that await characterisation. Several chapters of this book will help to take a step in this direction: consult those of Mieke Boon, Hubertus Nederbragt, Thomas Nickles, Andrew Pickering, Léna Soler, Emiliano Trizio and Frédéric Wieber. They will either point to schemes that can be seen as variations of the Wimsattian scheme: as refined or more complex versions of this scheme (see especially in this respect the more complex structural schemes designed by Trizio in Chapter 4, Fig. 4.4). Or they will exhibit alternative processes through which the status of 'solid' can come to be attributed to an element of scientific practices through time. In particular, in my contribution, I propose a complex architecture as a schematic reconstruction of the historical situation under scrutiny (see Chapter 10, Fig. 10.9). On the basis of this reconstruction, taken as having a general value beyond the particular case under study, I suggest regarding the Wimsattian scheme as a simple unit of analysis, an "elementary scheme" involved as a building block in more complex argumentative structures.

To see it as a building block could be one reason why, even if alternative schemes to the robustness one are involved in scientific practices, the robustness scheme is nevertheless especially fundamental. Other reasons can be mentioned. Trizio's article points to another very important one: even when the robustness scheme is not instantiated in a given episode of the history of science, it still works for practitioners as a regulative ideal or, as Trizio nicely expresses it, as a "methodological

[3] See Wimsatt (2007a), especially Chapter 8.

attractor". Indeed, the robustness scheme seems to incarnate the prototypical way through which practitioners of the empirical sciences think they can secure, if not justify, their experimental propositions, theories, models etc. We find incredibly numerous examples, in very different contexts of scientific practice, where scientists explicitly look for something akin to a robustness scheme and explicitly ask for such a scheme when it is not realized. As illustrated in Trizio's case study, reviewers of scientific papers often suspend publication of a submission until a supplementary, convergent derivation is obtained. Another common example occurs when scientists are convinced of a result *X*, and think their colleagues will be convinced too, because they think *X* is indeed involved in a robustness scheme (which also explains why scientific papers so often manifest robustness schemes, as Allamel-Raffin and Gangloff argue and illustrate in Chapter 7[4]). So, without doubt, the robustness scheme plays an effective and important role in scientific practices. Critics cannot reproach it for being an invention of the philosopher of science. In Stegenga's words, it is "an exceptionally important notion", "ubiquitous in science", and a "(trivially) important methodological strategy which scientists frequently use".

As a consequence, reflexive studies interested in scientific practices should analyze this role closely. Now, some of the existing studies seem to have followed what scientists themselves say uncritically and without philosophical arguments in their high valuation of robustness. Correlatively, some philosophers of science have invested very strong hope in robustness. In his chapter, Stegenga lists the many (themselves highly valued) epistemic tasks that robustness has been assumed to be able to achieve: to demarcate experimental artifacts from real phenomena and objective entities, to stop the experimenter's regress, to secure appropriate data selection, to help to recognize the best hypotheses (the most explanatory, empirically adequate, objective, otherwise promising), to provide a strong argument in favor of scientific realism, and more.

Given these hopes and the apparently pervasive involvement of the strategy of robustness in scientific practices, it is surprising that so few systematic historical, philosophical or sociological studies have been devoted to the topic: robustness remains poorly understood. The present book is intended to constitute a set of resources in order to improve this understanding. The contributors hope that their efforts will stimulate new investigations from others.

In the remaining part of this introduction, I will offer some systematic reflections that, although primarily directed toward robustness, are intended to be relevant to solidity more generally. So I see these reflections – which owe a great deal to the writings of two participants of this book, William Wimsatt and Andrew Pickering – as a contribution to the problem of solidity in science.

[4] The episode of the 'discovery of the weak neutral currents' on which I rely in my paper would also be a good example. See notably the quotations of scientists given in Schindler (201X).

1.2 Solidity, a Relational Status: Between Holism and Modularity

Solidity is a relational attribute: to give an account of the solidity of *X* amounts, inevitably, to invoke a multitude of elements *other than X*.

For example:

- To give an account of the *robustness* of $X =$ to invoke a *multitude* of partially independent derivations leading to *X*.
  Schematically and visually, the representation that first comes to mind is an arrows-node scheme, in which a series of upward arrows representing the derivations *converge* on the node *X* (see Fig. 1.1).

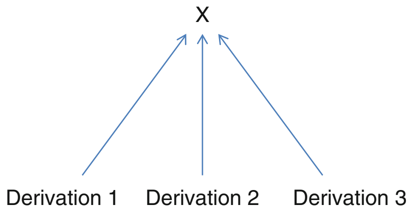

Fig. 1.1 The robustness scheme

- To give an account of the *GE* of $X =$ to invoke a *multitude* of practices in which *X* plays a quasi-foundational role (that is, for which *X* is required at a given historical moment, for which it appears impossible or very difficult to proceed without *X*).
  In terms of the arrows-node scheme, one is inclined to translate the situation by drawing a series of divergent upward arrows starting from the node *X* (see Fig. 1.2).

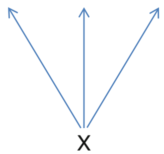

Fig. 1.2 The generative entrenchement scheme

Thus, solidity is not an intrinsic attribute of the *X* that is declared solid. The problem of solidity attributions has an irreducibly relational, holistic structure.

At the same time, we are not inclined to say that everything depends on everything else in science. Intuitively, we feel entitled to decompose the situation in terms of relatively autonomous modules, at least with respect to certain aims; and I think we must do justice to this intuition. The tension between holism and modularity is a crucial and difficult aspect of the problem of solidity in science. It will be encountered again below (Sections 1.6 and 1.7).

1.3 Counting and Weighing the Arrows of a Solidity Scheme

At first glance one expects that, all other things being equal, the solidity of a given node increases with the number of the associated arrows.

– With the number of arrows leading to the node, in the case of a robustness scheme;
– With the number of arrows starting from the node in the case of GE.

So analysts of a given historical scientific episode (among whom are scientists) must count the derivations: the higher the number of (independent) derivations, the more solid will be the *X* which lies at their intersection. But that is not all. Indeed, the different arrows are not necessarily on the same level. They do not always have the same strength. So the analyst must also weigh the derivations.

Let us begin with this second requirement and think about the robustness of an experimental fact in physics or in biology, for example – a kind of example that will occupy an important place in the contributions of this volume.

- While examining the various experiments that, through history, have contributed to the establishment of this fact as a fact, we are often led to introduce a hierarchy among the different experiments and to deem some of them more important, reliable and conclusive than others.

  Sometimes new derivations appear fragile because they are new, for instance if they involve innovative and relatively untried techniques or instruments that have to be checked. Catherine Allamel-Raffin and Jean-Luc Gangloff's chapter offers an illustration in the case of a new survey in astronomy. Sometimes new derivations are viewed as improvements on the old ones. The scientific episode in the field of quantum physics investigated by Catherine Dufour is a case in point. Here we find generations of successive experiments aiming at testing the same claim *X*, and each new experiment is seen as an improvement on previous ones with respect to this or that loophole in the experimental derivation.[5]

  Whatever the details of the particular historical configuration, it is clear that the different experimental derivations, past or present, that speak in favour of a given *X* in a given stage of scientific development are usually perceived as differing in strength. In her chapter, Dufour proposes a graphical representation of the situation by means of more or less large arrows according to the strength of the derivation.

[5] Admittedly, this is not a typical case of robustness, because as soon as real experiments are conceived as improvements on one and the same experiment, their independence is obviously discussable, and in any case they are not independent in the same sense as multiple experiments of different kinds. (For more details on this point, see below Section 1.9.2 the presentation of Dufour's contribution.) But as we will see, the independence is always problematic, and, anyway, the configuration involved in Dufour's case study is a widespread situation that must be considered in relation to the robustness issue.

- Moreover, the analysis of the solidity of an experimental fact should not just take into account *only* the experiments that have *supported* the result finally established as a fact, but *also* the possible experiments that have played *against* this result. For, in the controversial situations at least, the status of a solid experimental fact emerges through an assessment that is not based only on positive elements.

  Stegenga refers to this problem as “discordance”, which he divides in two types, “inconsistency” and “incongruity”. According to him, discordance is “ubiquitous” whereas concordance is not. Moreover, artificial concordance is often produced by retrospective readings of past science, through which all that is not congruent with the present science is ‘forgotten’ or dismissed. In the light of such considerations, philosophers of science should be more attentive to discordance in their analyses of robustness. Dufour’s study of quantum physics also raises the issue of discordant evidence and its consequences, in a situation where there exists a high multiplicity of concordant derivations against a unique discordant derivation which finally counts for nothing in the physics community.

  This difference between concordant and discordant arguments could be taken into account in more complex and enriched robustness schemes, and could be expressed graphically, by introducing, for instance, arrows with a (+) and arrows with a (–), or else by using different colours (with possibly a set of different shades indicating the relative strength of the supportive derivations on the one hand and of the ‘fragility-making’ ones on the other).
- Finally, the equation ‘robustness = invariance under multiple independent derivations’ only holds if the derivations involved are *genuine* derivations. In other words, it is related to a certain conception of what is an argument worthy of the name: it presupposes a set of norms about what is scientific/pseudo-scientific/non scientific.

  Indeed, imagine that the arrows of a robustness scheme represent derivations that have been rejected in the course of the history of science (for instance: the famous experiments invoked by Blondlot in favour of the existence of *N* rays). Or worse, imagine that the arrows symbolize derivations viewed as pseudo-derivations by any contemporaneous physicist. For instance: the *N* rays exist, because it is impossible that God, who is omnipotent, restricted himself to the narrow set of the rays so far discovered by physicists (electromagnetic rays, X rays and the like). Another amusing illustration is given in Hubertus Nederbragt’s contribution. Nederbragt mentions a situation imagined by Collins in which a scientific hypothesis under discussion would be tested, among other ways, by examining the entrails of a goat. This ‘method’ is indeed truly independent to the other scientific methods involved (such as the recourse to experiments), “But,” Nederbragt stresses, “here we reject immediately and with force the whole body of background knowledge of this method”.

  In cases of this kind, the derivations will only be considered as ‘derivations’ with inverted commas. Even if they are indeed independent of one another, they will not appear as potential arguments in favour of *X*, and the status of ‘robust’ will be refused to the node on which they converge.

These remarks point to the need for an analysis of what confers its strength (or degree of solidity) to the derivations involved in a robustness scheme. I will come back to this topic below in Section 1.6.2. In the meanwhile, we can conclude that the analysis of the solidity of an *X* will call for, together with the quantitative evaluation of the number of arrows, a prima facie more qualitative estimate of the strength of each arrow.[6] Not only must the arrows be counted, but moreover they must be weighed.

Actually the situation is even more complicated. Although at first glance an analyst's task of counting the derivations under which an *X* remains invariant seems straightforward, under examination it is not. Several reasons can be given for expecting that (and explaining why) different analysts (including scientists) interested in one and the same targeted scientific situation may represent it as involving different numbers of derivations in favour of *X*.

First, we can deduce from the previous reflections about the strength of a derivation that the two variables 'strength of derivations' and 'number of derivations' are not independent. As we have seen, a 'too weak' derivation, either positive or negative, will count for nothing. Clearly, generalizing, judgments of strength will have consequences for the number of derivations that will figure in the robustness scheme supposed to represent a given real situation.[7]

Second, there is the hard problem of assessing what should be counted as two truly different derivations or, rather, as two versions of *one and the same* derivation. This is what Jacob Stegenga calls the "individuation problem" (in his framework: for "multimodal evidence"; in my terminology: for multiple derivations): According to what criterion are we going to individuate a singular mode? The answer to the previous question is crucial, since it will determine the *number* of different modes

[6] Bayesians might contest the claim that the estimations of the arrow strengths are indeed "more qualitative", on the basis that they are able to quantify the weight of each arrow through numerical measures of evidential support. (I thank Stegenga to have drawn my attention to this point.) But, to my eyes, such kinds of attempts are at least useless with respect to the aim of characterizing real, ongoing scientific practices, and sometimes pernicious with respect to the realistic pretention of the numerical values they put forward and the algorithmic-transparent model of rationality they suggest. (see below, the end of Section 1.4, and Chapter 10, especially the conclusion.) So, at the end of the day, I think that robustness is indeed a qualitative notion (as is any human judgment), even if quantified modelisation can sometimes be clarifying. As Stegenga concludes, after having noticed that "Philosophers have long wished to quantify the degree of support that evidence provides to a hypothesis" and having regretted that in such context they "developed confirmation theory 'given a body of evidence *e*', without worrying about what constitutes a 'body of evidence'": "at best, the problem of discordance suggests that robustness is limited to a qualitative notion." I agree with Stegenga's conclusion, perhaps more strongly than he does himself, given his cautious formulation. In any case, the study of real, ongoing practices clearly shows that different practitioners very often weight differently the multiple derivations for and against a given hypothesis in a historical situation (see for example Dufour's paper as an illustration among others). So if weights could be meaningfully attributed to derivations, they would likely be different from one scientist to another one – which is of course not in the 'objectivist' spirit of the Bayesianist enterprise.

[7] For further developments on this issue, see also Section 1.5, third and fourth points.

that speak in favor of the hypothesis under discussion, that is, *how multiple* is the *multi*modal concordant favorable evidence for an *X*, and where lies the boundary between a *multi*modal configuration and a *mono*modal configuration.

1.4 Solidity, a Status That Comes in Degrees

'To be solid', as an attribute of an *X*, is not a binary property admitting a clear-cut application of the form 'yes or no'. *X* is not either solid or not solid: *X* is *more or less* solid. And, as a particular case, *X* is more or less robust, more or less GEed.

All authors who reflected on robustness have stressed this feature, among whom are Wimsatt himself, and, following him, many contributors to the present book. "Robustness may come gradually and in degrees", writes Nederbragt, for example, after having quoted Wimsatt.

But the next, more delicate and less explored question is: What does this 'more or less' depend on?

Let us consider the particular case of robustness. It already follows from the last section that in the assessment of the degree of robustness of an *X*, both a number of arrows and the strength of each arrow are involved. In addition, at least two other things must be taken into account.

- The *degree of independence* of the derivations, one from the others

  I will return to independence below, Section 1.8, but just to approach the idea intuitively at this point, imagine that the derivations turn out to be hardly independent, or not independent at all. For instance, imagine that they turn on a large number of hidden common assumptions. In such a case the convergence of their results will not provide additional support to the final result (additional with respect to the situation where only one of them would be available). The convergence would be artificially produced by the elements shared by the different derivations.
- The *more or less satisfying quality of the convergence* of the results derived from the different derivations

  At this level, what is needed is an analysis of the grounds on which it is possible to say that the multiple derivations lead to the *same* result. As several chapters of this book will clearly illustrate, the identity is rarely, if ever, immediately given as such. This is a very important point and, I think, one innovative contribution of this book. As far as I know, the reflections previously devoted to robustness topic did not really pay attention to it. As Trizio notes about experimental derivations in his case study in cell biology:

  > (…) different techniques often yield completely different experimental outputs. For instance, a colored film realized with *fluorescence microscopy* will have, at first sight, little in common with a black and white picture obtained by means of an electron microscope. In most cases, the identity is achieved only at the level of the judgments expressing the final results based on the interpretation of the experimental outputs.

Although not explicitly stressed by Allamel-Raffin and Gangloff, their case study offers an especially striking illustration of the idea that the multiple derivations of a robustness scheme lead to *strictly speaking different* derived results and not to *one and the same invariant* result. In some of the argumentative modules constitutive of the astrophysical paper they study, the robustness scheme extracted by Allamel-Raffin and Gangloff clearly involves *perceptually different* images obtained by different kinds of telescopes and different techniques of noise evaluation. Going from these different pictures to the conclusion that all of them converge (give the same information and hence can be considered as *one and the same* map), there is an inferential road and an act of synthesis: practitioners have to work on the initial images, they have to build ways (described in the chapter) to make them comparable and to be in a position to conclude that they converge. So what is at stake with robustness is constructing a convergence of a more or less good quality, rather than an identity that is simply given.

Mieke Boon also draws attention on this point and states it at a more general level. Contrasting "multiple means of determination" with repetitions of the same experiment, she stresses that the former

> usually does not produce the same results, at least not at the level of our observations or measurements. Yet, in his examples of multiple determination Wimsatt suggests that the point of it is producing the *same* results: "to detect the *same* property or entity", "to verify the *same* empirical relationships or generate the *same* phenomenon", etc. (Wimsatt 2007a, 45, MB's italics). This way of phrasing how multiple determination works suggests (. . .) that phenomena are like grains of sand on the beach. They are clearly identifiable objects in whatever circumstances: they remain as exactly the same identifiable entities whether on the beach, at the bottom of the sea, in the belly of a fish or in my shoes. (. . .) this is often not the case.

Stegenga, for his part, presents this problem as one of "the 'hard' problems of robustness", in the context of his discussion of discordant configurations, noting that these are often "incongruities" rather than clear inconsistencies.

> Evidence from different types of experiments, he emphases, is often written in different 'languages'. (. . .) To consider multimodal evidence as evidence for the same hypothesis requires more or less inference between evidential modes.

So, very often, the invariance, or rather the convergence, has to be built from the end-products of the multiple co-present derivations, and these end-products are initially different, strictly speaking. In my chapter, I analyze this building process for a case in which the operations of transformation are 'quasi invisible' and nevertheless – I argue – indeed still present, and I call the operations involved a (more or less creative according to the case) "calibrating re-description" of the end-products of the multiple arrows.

To sum up at this stage: *X* can be more or less robust, and this 'more or less' depends on at least four (not independent[8]) variables:

> *X* is (more or less) robust = a (greater or smaller) number of (more or less reliable) derivations that are (more or less) independent lead to a result (more or less close) to *X*.

This characterization leads to difficulties, the structure of which is admittedly familiar to philosophers of science at least since Thomas Kuhn.

How can we determine the 'threshold' beyond which *X* can be said to be solid, below which *X* must be considered as fragile? It is well-known that significant disagreements can arise at the level of the judgements of solidity/fragility of any element *X*, both within the members of the scientific community and within the members of the Science Studies community.

This is already the case if we consider in isolation each of the four degree-judgments involved in the robustness equation above. And it is a fortiori the case for the global judgment in terms of the degree of robustness, which, in addition, involves a balance judgment. The latter description is, in my view, an analytical decomposition and reconstruction of the situation, provided for the purpose of clarification. In practice it is more than dubious that scientists proceed by composition from the parts to the whole, that is, first by assessing the strength, independence, degree of convergence, and number of convergent derivations, and, second, by constructing, on the basis of these 'atomic' judgments, a 'molecular' judgment about the resulting robustness of *X*.

Moreover, the global judgment about the robustness of an *X* can vary, 'all other things being equal', with contextual conditions 'external' to the robustness scheme itself. For example, and as illustrated by Trizio's case study, if an experimental result *X* under discussion is, according to the theories admitted at the time, highly implausible, then a higher number of convergent experimental derivations will be demanded to in order to conclude that *X* is sufficiently robust – higher than if this *X* were expected according to these theories. Or, again, and as illustrated by Dufour's case, if *X* appears as an especially crucial and important proposition, even a very high number of concordant experimental derivations will seem still insufficient to some scientists.

This is not to say that, as philosophers of science, we cannot try to propose criteria, or at least to reflect on what could be the criteria, to evaluate what is a sufficiently high number of sufficiently independent derivations which show, taken one by one, a sufficient degree of strength and, taken all together, a sufficiently good degree of convergence, and on this basis, criteria to assign a sufficient degree of robustness. In this vein Stegenga introduces what he calls "the amalgamation problem", that

[8] In Section 1.8, we will see that the variable 'number of derivations' is no more independent of the variable 'degree of independence' than of the variable 'strength of the derivation'.

is, the problem of "*how* multimodal evidence should be assessed and combined to provide systematic constraint on our belief in a hypothesis" *X*, and he looks for an "amalgamation function" to do the job. But his analysis points to deep difficulties associated with the task.

So at the level of all of the degree-judgments involved, we face the difficulty, not to say the impossibility, of specifying universally compelling criteria and of pointing at a universally compelling threshold beyond which it would be de jure irrational not to uphold the solidity of *X*. This difficulty becomes particularly tangible, and its epistemological implications become clearly visible, when we study moments of the history of science during which important controversies about solidity judgements took place among the practitioners themselves. However, this difficulty is in principle also present for episodes in which a good consensus exists among the members of a scientific community. And it is a difficulty with respect to which a philosopher of science willing to analyze the solidity of what is taken as a scientific accomplishment has, sooner or later, to take sides, even if it is by entrenching himself behind the practitioners' judgements.

1.5 Arrows-Node Schemes of Solidity and Scientific Practices

The volume title refers to the practice turn. What then is the relation between a solidity scheme – such as the one of robustness or generative entrenchment as defined above – and the level of scientific practices?

Solidity schemes can be said to be *emergent* with respect to scientific practices, in the following sense:

- The scheme aims at grasping what has emerged during certain 'conclusive moments', at a given point of the historical process.
- The arrows and the nodes that constitute the scheme belong, therefore, to the level of the 'finished' products of science, rather than to the level of the ongoing process of scientific investigation;
- The arrows and the nodes refer most of the time to what appears in scientific publications, rather than to what is actually done in the laboratories.

This doesn't diminish the interest and relevance of these schemes, for the vast majority of the scientists themselves rely on what appears in the publications in order to judge the solidity of a procedure or of a scientific result. But this makes us suspect that the robustness and GE schemes do not exhaust the analysis of the constitution of solidity. I will shortly get back to this when I address the issue of the independence of the derivations (Section 1.8).

Meanwhile, I would like to specify further in what way the solidity schemes pictured by means of the arrows-nodes representation belong to an emergent level with respect to the one of science in action and in real time.

I will specify that point by relying once more on the case of robustness.

- First, the robustness scheme already presupposes the stabilization of the 'derivation-result' connection. Whereas in real and developing laboratory practices, this connection emerges from a more or less delicate and more or less problematic trial and error process, the outcome of which is of course not determined in advance for practitioners.
- Second, the scheme offers a panoramic description that treats each derivation as an unproblematically individuated entity and as a black box. Whereas with reference to laboratory practices, a derivation involves multiple sequences of actions, whose individuation and re-individuation are not obvious and may be problematic. Moreover, even once these sequences of action have been conceptually grasped as *one* sufficiently well-defined *kind* of procedure symbolized and black-boxed as an arrow, this arrow actually refers to a complex reality combining multiple components of heterogeneous types (material-conceptual-functional objects such as instrumental devices, practical skills required for the successful use of the instrumental devices, various kinds of operations and reasoning. . .).
- Third, and as already stressed Section 1.3, the scheme only takes into account the supportive arrows (the 'confirmations'). It says nothing about the argumentative lines that could weaken the node *X*, if not refute *X*. In other words, the scheme holds after an assessment of the importance of the negative arguments possibly involved in the historical situation. Such an assessment can be either the conclusion of a deliberate and systematic discussion, or – maybe more often – an intuitive judgment associated with implicit reasons. But in any case, the robustness scheme, which does not even mention, and thus ignores, the negative arguments, presupposes that these negative arguments are not significant or of very weak weight.[9]
- Fourth and finally, even with respect to the supportive elements alone, the scheme presupposes a preliminary choice, related to the more or less significant character of the available positive arguments. As already stressed and illustrated in Section 1.3 above, only those deemed to be the most significant ones will appear as arrows in a robustness scheme.

Clearly, a solidity scheme is an idealized representation of a real scientific configuration, which depends on and reflects multiple judgments and decisions on the side of the analyst who proposes it.

---

[9] This point is closely related to what has been developed above, Section 1.3, about the issue of the weights associated with the derivations. An alleged negative argument once proposed in the history of science against an *X* but taken by a given analyst to be very weak if not totally dismissed as an argument (equated with something that counts for nothing, or if one prefers, associated with a weight equal to zero), will *not appear at all* in the robustness scheme that this analysis proposes as a reconstitution of the historical situation.

1.6 About the Nature of the *X* Appearing in the Judgment '*X* Is Solid'

1.6.1 Solidity of the Nodes

I would like, now, to say a few words about the nature of the *X* that could appear in the judgment '*X* is solid'.

Actually, this *X* can be of quite various kinds. This variety will be exemplified by the diverse types of *X* candidates for robustness attributions that are involved in the different chapters of the book. Several contributions also directly address the issue of this variety and propose typologies or classifications. For example, Boon's chapter attempts to list systematically and to categorize the multiplicity of heterogeneous 'things' that are commonly called "robust" and to explain the different senses involved.

A. Under the influence of a long tradition, philosophers of science first think of *X* as a theoretical or experimental result that could appear in a descriptive sentence about the natural world. More generally, they think to *X* as a proposition (or a set of propositions).

   When ***X* = a propositional entity**, '*X* is solid' (typically) means:

   X is trustworthy, reliable, effective, believable, if not true;
   The entities and processes referred to by the proposition exist, are objective, are real.

   For instance, *X* can be a theoretical hypothesis, a whole theory, a theoretical analysis or explanation, an experimental fact, a scientific model, a certain belief, and so on. Many chapters of this volume will consider the solidity of this kind of *X*, as the examples invoked up to now in this introduction already show. In most of these examples, *X* was a proposition, and, in addition, all those examples came from the empirical sciences. But the mathematical sciences are also considered in this volume. In Chapter 8, Ralf Krömer examines the peculiar case of the robustness of mathematical propositions (and more generally if, and to what extent, the robustness scheme plays a role in mathematical practices).

B. Besides propositions, we can also think of other types of scientific results that can take the place of nodes in a solidity scheme.

   For instance, ***X* can be a scientific image** in a broad sense of the term "image" (a photograph, a picture, a map, a pictorial scheme[10]...). In their chapter, Allamel-Raffin and Gangloff analyze a case of this kind in the field of astronomy.

[10] Some might claim that images or maps are reducible to complex networks of propositions and hence do not correspond to a *different type* of node than in the preceding case. This might be a controversial point. In their paper, Allamel-Raffin and Gangloff argue that images are not reducible to propositions and that they play a specific role in scientific argumentation. I am inclined to think they are right. Anyway, it is not my problem here to discuss this point. My intention is only to illustrate the diversity of the scientific items that can possibly constitute a node in a solidity scheme.

C. We can also envisage the case in which the node X of a robustness scheme refers to a measuring instrument, or more generally to any technological device.

When $\boldsymbol{X}$ **= a measuring instrument**, 'X is solid' (typically) means:

X fulfills, reliably and durably, the function for which it has been conceived (the measurement of this or that quantity or phenomenon);

The instrumental outputs of X provide information about the 'true values' of the variables, about characteristics actually possessed by the measured objects.

At this level and more generally, the task is to investigate what solidity may mean when $X =$ a technological means, and how exactly the solidity of these hybrid entities is constituted, given that they are, at the same time, material, conceptual, and intentional, that is, designed in the intention to fulfill definite practical functions with respect to human life.

In Chapter 13, Andrew Pickering reflects on the solidity of technological material achievements (what he calls "free-standing machines and instruments"). Faithful to the fundamental intuition that underlies all his work, he focuses on the level of the most concrete actions and material aspects of science, holding that "material performance and agency" is "the place to start in thinking about the robustness of science".

The contribution of Nickles also touches on the solidity of technological achievements, but here the X under discussion is not just a simple localized object such as a measuring instrument; it is, rather, a highly complex "epistemic system" which involves, among other dimensions, technological features.

D. We can add this kind of complex X to our repertory. True, contrary to what the title of the present section suggests, we are not intuitively inclined, initially, to see this kind of X (a complex epistemic *system*) as a node. But this could be said as well about the X considered just above section C, for a measuring instrument can itself be viewed as a (more or less) complex system and represented as an extended network constituted by multiple nodes and arrows. And after all, any complex system can be seen, at a certain scale, as one unitary entity, and its solidity can then be discussed as such. Anyway, even if the kind of X in question is not at its ideal place in the present section, the discussion of the solidity of $X =$ a complex hybrid system is an important aspect of the solidity problem.

When $\boldsymbol{X}$ **= a complex hybrid system**, 'X is solid' can mean very different things, beyond the idea of its ability to fulfill, reliably and durably, the function(s) for which it has been designed. It depends on what the system is made of, on its intended use and on what the users want to avoid. Admittedly, the solidity will be differently conceived if the complex system is an epistemic system like a scientific paradigm, an industrial system like a nuclear power plant or an information system like the Internet. But at a general level we can nevertheless stress, following Nickles in Chapter 14, that for complex hybrid systems, to specify what solidity is, requires us to specify the *kinds* of anticipated threats or failures we want to avoid, and to *relativize* solidity *to* these kinds. A complex system is never solid in all respects. It may be solid in certain respects but fragile in others. Moreover, as Nickles argues in his chapter, attempts to increase the solidity in one direction often, as far as we can know, create new, unpredicted fragilities, sometimes worse than the ones eliminated or attenuated by previous design modifications. In these cases a "robustness-fragility tradeoff" is inevitable and we must renounce once and for all the optimistic idea of a "cumulative fragility-reduction".

With respect to complex systems, more especially technical systems, Nickles, in his chapter, refers to a useful distinction between two kinds of robustness (or rather, solidity, according to my terminological decisions) that we can fruitfully add to our toolbox in the context of this introduction. The first is "simple robustness", referring to a situation in which a system is robust because it is made of a few very reliable components (the robustness of the totality is the sum of the robustness of the parts). The second is "complex robustness", related to a situation in which the robustness is an emergent property of a complex system made of multiple 'sloppy' and cheap *but redundant* parts and of diverse processes of control.

E. Finally, outside of the field of studies devoted to scientific achievements, robustness is also currently applied by historians and philosophers of evolutionary biology to living beings 'designed by nature' (by reference to and by analogy with technological devices designed by humans), or more generally to aspects of the natural (possibly nonliving) world, such as physical phenomena or the behavior of physical objects. This dimension of robustness will not be systematically investigated in the present volume which is primarily interested in the robustness *of scientific achievements*. Nevertheless, Wimsatt's contribution will consider this kind of "material robustness" (as he calls it, by contrast to "inferential robustness"), and will illustrate it through the example of the robustness of some phenotypic properties required in order for the sexual reproduction to be possible.

1.6.2 Solidity of the Arrows

The previous examples were examples of an *X* that would appear *as a node* in a solidity scheme. Now I would like to say a few words about the solidity of the *arrows* of the scheme.

I will consider again the case of robustness: the case of an arrow that denotes a procedure (a derivational means, whatever it is) aimed at establishing the robustness of *whatever type* of scientific result. At this level, the typical examples are:

- A theoretical argument in the empirical sciences;
- A mathematical argument;

- A protocol of any kind, for instance:
  - A technique (of fabrication, of production of samples, of preparation...)
  - An experimental procedure, with its characteristic set of instrumental apparatuses and techniques...

When X = **a procedure-arrow**, 'X is solid' can be further characterized as: X is reliable, trustworthy, efficient, stable.[11]

Several chapters of this book deal with the issue of the solidity of a procedure in an indirect or peripheral manner, but this question is the central topic of Frédéric Wieber's contribution (Chapter 11), which describes and discusses how a modeling procedure developed in the 1960s and 1970s in the field of biochemistry has been progressively taken as a solid procedure. Dufour also directly addresses this issue, through the characterization of two different "loopholes" in the available derivations in favor of the violation of Bell's inequalities, and insistence on the fact that practitioners assess the consequences of these loopholes differently. This is also the topic of my own chapter, where I open the black box of an experimental derivation in favor of the existence of weak neutral currents in the 70s, and give a structural characterization and a visual representation of this derivation, on the basis of which I discuss what the solidity of a derivation is made of.

If we ask: 'What types of reasons can be invoked to support the reliability of X = a procedure?', the answer can be explored along two paths and divided in two parts.

- A first part of the answer doubtlessly lies in a sort of reversed formulation of the Wimsattian robustness scheme: the solidity of a procedure will increase with its being involved in the derivation of the greatest possible number of independent results already established as solid (say the $R$is). For instance, the reliability of an experimental procedure will increase with the number of already-robust results that it yields.

  Does this first part of the answer lead us in a circle?

  Yes, no doubt. But the circle is not necessarily vicious. On the face of it, we are tempted to avoid the vicious circle by pointing out that the results and methods that play the role of 'what establishes solidity' on the one side, and the role of 'what acquires solidity' on the other side, are not the same. The

[11] Note that the same terms were already employed in the previously considered case, 'X = a propositional entity'. Actually, these terms, and especially the reliability vocabulary, have a very broad scope and can be applied to propositional results as well as to procedures. (In this respect, the lexicon of 'reliable' could have been a good alternative to the vocabulary of 'solidity'). But note also that the reverse does not hold: all the terms used to name the solidity of a *propositional* result do not automatically apply to *procedural* achievements. In particular, talk of 'truth', 'objectivity' and what is 'real' are restricted to propositional entities. They are not, in ordinary usage at least, employed to pick out procedures, methods, derivations and the like. These differences at the level of the vocabularies reflect a difference of epistemic kinds.

already-solid methods act like springboards to establish the solidity of *new results*, and the already-solid results act like springboards to establish the solidity of *new methods*. We find, therefore, rather than a circle, a helix.

These reflections lead to several suggestions.

First, they clearly show that a solidity scheme only holds with respect to certain historical coordinates, and, therefore, that any solidity scheme should specify these coordinates. Indeed, these coordinates will determine what can be taken as already solid – what works as an already established ingredient – and can therefore serve as a source of solidity for the *X* under scrutiny.

Second, even if we believe that this helix is not vicious, we are pushed in an endless journey back trough the axis of the time of the history. For the 'what has been *newly* established as solid' always refers to 'what has been established as solid *before*', and so forth. We have a temporal regression without an end.

This can make us suspect that the historical order is far from being indifferent. . . That science might be path-dependent in a highly constitutive way. It raises the problem of knowing whether we can leave the historical order behind, by considering, as is almost always done, that at least some of our robust scientific results enjoy autonomy with respect to this order, and were, in a sense, inevitable independently of this order. This is a delicate question. The epistemological implications are the degree of contingency associated with what is taken as a scientific result, and the plausibility of a science that would be, *at the same time*, *as solid as our* science, *but* coordinated with an *irreconcilable* ontology – an ontology either contradictory, or, more plausibly, incommensurable to ours. This is the contingency issue, to which Andrew Pickering has made an original, if controversial, contribution, and to which I shall return at the end of this introduction.

- The answer that has just been given to the question of what provides its solidity to a derivation leaves us unsatisfied. It seems to be only a part of the answer. Indeed, intuitively, this answer appears too 'extrinsic' with respect to the procedure *X* whose solidity is under discussion. For it depends on historical circumstances external to *X*, that is, to the *empirically contingent* development of a set of *other* procedures involved in the establishment of the solidity of the already-established-as-solid *R*is.

  Intuitively, we would like to link the solidity of a procedure to some of its 'internal' properties, to more 'intrinsic' features. This would be the second part of the answer to our question. In my chapter, I will examine the nature and the role of these possible intrinsic features that are expected to give its solidity to a procedure 'from the inside'. But to anticipate, I can emphasize the fact that these features – that one may be inclined to consider, at first sight, as 'intrinsic' – can be said to give its solidity to a procedure only *by referring to many other things besides themselves.*

On the whole, we have the quite vertiginous impression that every time we ask about the solidity of an ingredient of the robustness scheme, we start a cascade involving more and more ingredients, horizontally in the synchronic space, and vertically along the diachronic axis.

I will now have a closer look at what there is behind this impression. This will lead us back to the tension between holism and modularity introduced above in Section 1.2.

1.7 From the Pyramidal-Foundational Model to the Holistic-Symbiotic Model, and Back

In order to understand better the 'mechanism' that underlies the attributions of solidity, I will discuss the simple case of a four ingredients robustness scheme in which three arrows converge on a node *R*, and I will do the following exercise: to vary the 'solidity values' (an expression framed on the model of the widespread expression 'truth value') associated to the different ingredients of the scheme, and to examine the various resulting readings.

1.7.1 A Thought Experiment Playing with the Solidity Values of the Elements of a Robustness Scheme

A. Let us first imagine that, at a given moment of the history of science, the three kinds of methods involved in each of the three derivations are – independently and before they come to be seen as able to derive the result *R* and as a supportive argument in favor of *R* – already standard and well-mastered methods, and thus methods seen as reliable and solid. Let us assume moreover that the three arrow-methods are of *equal* solidity.[12]

   On the basis of such 'initial'[13] solidity values, the scheme of robustness should be read as follows: the result *R* becomes robust (or: acquires a strong

[12] This is of course just a thought experiment in order to show how the solidity scheme works, not a positive claim about the possibility of 'measuring' the solidity values that in fact hold or should have held in a given empirical situation. As already suggested in note 6, this possibility is, to say the least, highly problematic.

[13] The quotes are intended to prevent the reader from equating this 'initial' to a given moment of the history of science at which one would have succeeded in measuring the actual, objective solidity values of some existing scientific derivations. This is not at all the sense of the present exercise. The 'initial' and the 'final' name two *logical* moments of the proposed thought experiment. The thought experiment *imposes by fiat* such and such solidity values to certain elements of the robustness scheme (which are the 'initial' values) and then discusses, on this basis, what consequences result 'next' (i.e., what are the 'final' solidity values and how they result from the 'initial' state). Of course, this thought experiment is intended to help us to understand what happens in real historical situations. But this does not imply that in real historical situations anyone is able to associate explicit objective measures to scientific derivations. The relation with the history of science, and the way the thought experiment is used to understand real situations, will be sketched below Sections 1.7.2 and 1.7.3.

degree of robustness), because three independent methods converge on it. Here, the 'flux of robustness' is *oriented* from some ingredients of the scheme (the three derivations taken altogether) to others (the result *R*): namely *from* the more robust ingredients *to* the less robust ones. Intuitively, one is easily inclined to characterize such a situation by means of a pyramidal-hierarchic (down-up) model, if not by means of a foundational one: a model according to which an 'intrinsically solid' basis works as a foundation for the edification of upper, 'intrinsically more fragile' floors.

B. But let us now modify, in the preceding scenario, the 'initial' solidity value of one of the arrow-methods. Imagine that one of the three methods is controversial, that it has a low degree of solidity. In such a configuration, the robustness scheme will lead to a reading notably different than the one in the previous 'democratic' scenario (same weight attributed to each arrow).

   It will be read as follows: two equally solid derivations (based on solid methods) lead to the result *R*, thus this result acquires a certain degree of robustness. In parallel, a method so far considered as fragile leads to the same result *R*. Thus, the method in question acquires solidity. Therefore, on the whole, we can say that the result *R* is at the point of convergence of *not just two* solid derivations, *but three* solid derivations, and thus, the degree of robustness of *R* still increases (compared with the configuration in which *only two* arrows converged on it).

   The description just proposed to describe this second scenario appears a little bit artificial and unsatisfactory. This is mainly because a verbal presentation is ineluctably sequential, joined to the circumstance that, for the sake of didactic purposes, we are led to decompose the situation into a maximum of 'partial fluxes' of robustness. In such a verbal presentation, we cannot but consider *one after the other*, and hence in a certain order, the multiple partial fluxes of robustness directed from some ingredients to some others. Whereas, in fact, the robustness attributions result from a global, 'instantaneous' equilibrium, from an unordered mutual reinforcement.

C. Actually, the most suitable configuration to show that the attributions of solidity result from a global equilibrium is the configuration in which all the ingredients of the scheme are 'initially', when considered in isolation from one another, all taken to be fragile (or associated with a weak degree of robustness).

In that case, the scheme will be read as follows: the robustness of the result *R*, the solidity of the first derivation, the solidity of the second derivation, and the solidity of the third derivation, are each *mutually and equally* reinforced, due to the global configuration in which they are involved. *Although 'at the start', none of the ingredients were solid, all of them gain solidity from their conjunction in a network.*

In such a configuration, the 'solidity fluxes' are no longer unidirectionally oriented gradients: they are uniformly distributed. The reinforcements act in both ways and in a symmetrical manner between all the elements, a situation that we can represent by means of double arrows. Each ingredient equally contributes to the solidity of all the others.

In such a configuration, one is no longer inclined, intuitively, to appeal to a hierarchic-foundational model. Rather, one is inclined to say that the solidity is co-constituted, co-stabilized. One is inclined to describe this situation by referring to a network structure, to a holistic equilibrium, to reciprocal stabilizations, or again – to borrow a term often used by Pickering in his analysis of scientific practices – to a "symbiosis".[14]

Here not only are we back to the relational nature of solidity attributions noted above, in Section 1.2, but we can also go a step further. Relying on the previous analysis, and taking into account the fact that each element of the network owes its strength to its connection with the others, we are led to conclude that the *X* involved in the judgment '*X* is solid' can be identified – and perhaps is more adequately identified – with a more or less complex structure or network or symbiosis, rather than with one particular ingredient of this structure (as is commonly done and as I have treated it in this introduction up to this point). In his chapter, Nickles will approach robustness by taking up this perspective, that is, from the angle of the general characteristics that complex *systems* might or should satisfy in order to be solid in this or that desired respect.

1.7.2 What Hides the Holistic-Symbiotic Working of a Robustness Scheme in a Given Historical Configuration

When we analyze a particular case, what is it that hides the holistic-symbiotic character, or even leads to reject as invalid the holistic-symbiotic model and favors, instead, a hierarchic-foundational reading? It is the operation that, 'initially', remains fixed, and assumes as unproblematic the solidity of some of the ingredients of the scheme considered in isolation.

What can we say about this operation?

First, that it obviously reflects what is the case in the history of science and in real practices. Indeed, each practitioner, at a given period of the research, does not place on the same level all the ingredients involved in the developmental stage in question.

Second, we should not deplore such a situation, since it seems to be a necessary condition of the possibility of our science. If the attributions of solidity were all the time and all together questioned, practitioners could not rely on anything and science could no longer progress – or in any case, would not look like *our* science. We are led here to a variant of the famous metaphor of Neurath's boat.

Third, the question nevertheless arises of the origin of the attributions of solidity that are 'initially' associated to the multiple ingredients of the robustness scheme.

[14] Pickering's symbioses apply to real time dynamic scientific practices and their various (material, intellectual and social) ingredients, rather than to elements of an idealized and 'synchronic' robustness scheme as is the case in the reasoning just above. For more about the symbiotic conception of scientific development and references, see Section 1.7.3.

And to this question, the answer seems to be: the initial solidity of an ingredient considered in isolation is related to its past history, and more precisely, to the way this ingredient has been, in the past, involved in other holistic equilibriums or symbioses. We are back to the open-ended helical process we met before. We have the impression that the holistic modules represented in the plane of the sheet of paper must always be connected, 'in depth' (in a perpendicular plane representing the temporal dimension), to an undefined multiplicity of structurally analogous holistic modules.

1.7.3 Structural Homologies and Substantial Differences Between the Robustness Analysis of Real-Time Scientific Practices and Retrospective Consideration of Past Science

What has just been said about the holistic-symbiotic way of functioning applies, as I stressed before in Section 1.5, to a level which is emergent with respect to scientific practices. In Section 1.5, I specified in what sense and for what reasons the robustness scheme is emergent. We can now reformulate the point in terms of holistic units, now seeing any solidity scheme as a holistic unit.

A solidity scheme reflects and carries some options about the content and the extension of the holistic-symbiotic units within which, and at the scale of which, the interactive stabilizations are supposed to apply. In other words, it presupposes options about the content-extension-frontiers of the network on the basis of which it is relevant to think about the equilibrium of the solidity distributions. That is, the analyst cuts through the infinitely rich historical reality where, strictly speaking, everything could be thought to be linked with everything in a certain respect; and he extracts a holistic module considered as sufficiently autonomous with respect to the aim of solidity analysis. In other words, he assumes *local* holism[15] and it remains for him to decide about the 'size' of the local. He has to face the tension between holism and modularity and to overcome it in one way or another.

That said, *although emergent*, the analysis proposed above concerning the way the holistic modules function, is, I believe, *still relevant*, in its general principle, for understanding what is at stake in *real practices* such as laboratory practices. This is so, since something like the same structural scheme is involved in both cases: at the level of the emergent scheme as well as at the level of the ongoing practices, we find interactive co-stabilizations. When we consider science from the standpoint of

[15] Thomas Kuhn introduces the idea of "local holism" in a 1983 paper in which he tries to understand the incommensurability of scientific theories in terms of the (locally) different linguistic structures of these theories (On Kuhn, robustness and incommensurability, understood in a structural framework inspired by network theories, see Nickles' paper, Chapter 14.) To decide how local is this local holism (and hence how 'extended' or 'spread' is the incommensurability), or to decide the size of the network relevant to robustness assessments, is the same kind of problem and leads us to back to the tension between holism and modularity stressed Section 1.2.

practices, we can re-describe it as an attempt to obtain good "symbioses"[16] between the ingredients involved in these practices. When the attempt is successful, what is obtained is a modular (more or less extended) holistic unit endowed with a certain quality of stability and autonomy. Ian Hacking once described it as a "closed system" and a "self-vindicating structure" (Hacking 1992, 30), Andrew Pickering as a "self-containing, self-referential package" (Pickering 1984, 411). It is such a unit which can be represented at an emergent level, through considerable simplification and inevitable distortions, by an arrow-node scheme of solidity.

Now, if there are structural homologies between what holds in real historical situations and what holds at the level of the very simplified representation of solidity schemes, there also exist substantial differences.

A first striking difference, to which an allusion has already been made at the very beginning of this introduction, lies in the *nature of the ingredients* involved in the inter-stabilizations under scrutiny. When we consider science in action, for example laboratory practices, the interactive stabilizations or good symbioses that eventually emerge arise, as the practice turn showed, among ingredients *of various heterogeneous types*, possibly including know-how and professional skills, local norms and standards guiding the production of results, instrumental and material resources, geometry of the laboratory and short-term concrete feasibility, if not available institutional organization, the power of convincing peers, decision-makers, sponsors, and so forth.

If we want to continue to use an arrow-node diagram in order to represent living practices and the process of constitution of solidity in these practices, we will have to distinguish and to define the possible kinds of ingredients potentially involved and to find a specific representation for each kind.

On this path, the task and the difficulty are:

- To specify and classify the items that may constitute the robust scientific symbioses (taxonomy of the constituents that are likely to intervene)[17]
- And at the same time, to specify the nature of the 'glue' that may be able to 'hold together' the various ingredients of scientific practices (typology of the relations that are likely to contribute to a solid symbiosis). We need to find a way to think

[16] Pickering (1984, 1995). Hacking sometimes also uses this terminology (see Hacking 1992), but more 'in passing'. In Pickering's writings, the idea of a symbiosis is more developed and is a central conceptual tool.

[17] This is actually a very difficult task. Hacking made an attempt in this direction in Hacking (1992). He distinguishes three main categories ("things", "ideas" and "marks") and fifteen elements distributed on them (See the quotation note 21 below for examples of these elements. See also Mieke Boon's paper, Section 1.2, for a presentation and brief discussion of Hacking's typology). More generally, and as already noted, Boon's paper itself aims at building a typology of the ingredients involved in scientific practices that might be candidate to robustness attributions. Further work on this issue is, in my opinion, strongly needed. Clearly, the general – highly criticized but still in use – distinctions such as cognitive/social and internal/external factors are completely non-operative for the analysis of detailed case studies.

about the nature of this 'global hanging together' on the basis of something different from the classic idea of logical coherence. We need to develop something akin to what Hacking called an "enlarged coherence" between "thoughts, actions, materials and marks" (Hacking 1992, 58).[18]

A second important difference between a characterization at the emergent level of solidity schemes and a characterization at the level of real practices[19] is that, in the latter case, we must take into account a sense in which at least some of the ingredients involved in real time practices are in a relation of co-*maturation* (this is one important aspect that the idea of a symbiosis is intended to convey).

This means that, at a given moment of the history of science, several coexisting traditions (typically, some theoretical and some experimental) reciprocally feed the others with relevant and interesting subjects of research and mutually influence the investigation of the others in certain directions, thus at the same time moving them away from different possible directions. Each pole favours what is 'in phase' with it within the other (where 'in phase' means at least 'provides relevant exploitable elements to it', if not 'gives positive support to it'). Such a picture motivates the symbiosis metaphor: each pole gives life to, or sustains, the survival of what is 'in phase' with it. Correlatively, each pole neglects or dismisses what is not 'in phase', thus contributing to its extinction. When, as a result of this process, a better symbiosis than before is achieved at a point of the history of science – as sometimes happens –, that is, when practitioners feel that more things than before nicely fit together or are more strongly connected, this in turn conditions what happens next – what is taken as an interesting question, as an already solid and as a still fragile ingredient, etc. – and so on from point to point along the temporal axis of the history of science.

As a consequence, an arrows-nodes scheme designed to be a model of real scientific practices should represent the various equilibriums *in 'real time'* and show how they have been restructured along the historical path. This means that each arrow-node configuration should only involve *contemporaneous* ingredients – rather than juxtaposing in the same space, as is often the case in an *emergent* scheme, derivations that have been stabilized at very distant moments of the history of science. More precisely: if the scheme juxtaposes derivations that have been stabilized at distant moments, the arrows should represent, not these derivations as they appeared to practitioners at the time they had first been stabilized, but – and this is almost always very different – these derivations *as they appear to scientists at one and the same later moment of the history of science* (which coincides either with the period

[18] According to the kinds of 'ingredients' and kinds of 'glue' one is ready to associate to a good/better scientific symbiosis, the idea of a symbiosis can carry different philosophical implications, and in particular, stronger or weaker relativistic implications. See Soler (2008a, 2006b).

[19] The degree of complexity and the opacity of the configurations could also be mentioned here, but I will leave these aspects aside.

of the oldest derivation involved in the scheme or with a period posterior to all the derivations involved).

In this book, several authors – usually inspired by the work of Pickering and by the 1992 paper of Hacking on the self-vindication of science – develop their reflections on robustness (and the dynamical arising of robustness through time in the history of science) in terms of mutual adjustments and interactive stabilizations.

First of all, Pickering himself does this, of course. In his framework, scientific development appears as a "dance of agency" with passive phases where the "otherness" of the world manifests itself through material performances, and active phases where human beings try multiple accommodations and adjustments. After many iterations of such phases, sometimes a point is reached where practitioners feel that a good "interactive stabilisation" (or good symbiosis) has been achieved: all the pieces of knowledge nicely fit one with the others and some pieces of this system nicely fit with the material performances of available material means, so that at the end of the day, all items reinforce one another. At this point the dance extinguishes itself and a duality is produced between the human and the non human (natural or technical) world. At this point some nonhuman ingredients appear to be solid scientific results (supported by other ingredients of the structure) and can be published. As I understand Pickering's picture, on the epistemological side,[20] the robustness of an ingredient of scientific practice is due to the good 'hanging together' of the multiple conceptual and material items involved in a given stage of scientific development.

Boon, following Hacking, also endorses the conception that what practitioners of science strive to achieve is the mutual adjustment and the co-stabilization of multiple elements of different kinds. All the ingredients involved in such self-vindicating structures deserve, according to Boon, the status of scientific results – not just facts and theories as in traditional accounts. As a consequence, many different sorts of scientific results ("data, physical phenomena, instruments, scientific methods and different kinds of scientific knowledge") can be said to be robust, in different senses of the word. In this framework, Boon's chapter aims at specifying how the different kinds of robustness attributions are acquired and what are the interrelations between them.

The same kind of framework underlies my own article. Indeed, my analysis of an experimental derivation makes this derivation appear, at the end of the day, as a complex architecture and a global equilibrium between a great multiplicity of elements that fit altogether.

I think something of the same kind is involved in Nickles' reflection on complex epistemic systems, despite the fact that his reflection is framed in different categories and refers to a different background literature (network theory, theories of complex systems and risk analysis) than the works just mentioned.

[20] In his contribution to this book, Pickering also considers (and primarily focuses on) the *ontological* side of robustness (see below, Section 1.9.5).

1.8 Independent Derivations . . . In What Sense?

The last point I would like to examine in this introduction concerns the independence of the derivations mentioned in a robustness scheme. Several contributions to the present volume encounter this issue at one point or another, and three of them, namely Nederbragt's, Stegenga's and Trizio's, address it extensively.

This is a delicate but highly important issue for at least two reasons: first because genuine independence seems to be a condition of possibility of genuine robustness; second because the independence of convergent derivations is rarely explicitly analyzed by scientific practitioners. As Nederbragt stresses in his contribution, if, "in daily practice scientists discuss experimental methods and their results", in these, "evaluations independence is (. . .) a hidden criterium, applied intuitively."

In order for the *N* derivations represented in the robustness scheme to be stronger than just one of them (or stronger than a number smaller than *N* of them), there have to be, among them, some differences *that make a difference*. The plurality must be real and not just an illusion. The derivations shouldn't, after scrutiny, turn out to be, in some sense, derivable from one another. If this were the case, the 'at first sight multiple' *N* derivations would reduce, under examination, to a number of derivations inferior to *N*, and possibly to one unique derivation. Clearly, the 'degree of independence' and the 'true number' of derivations are not two independent variables.

In order to describe this requirement, Wimsatt resorts to the vocabulary of (partial) independence: The derivations must be, two by two, at least partially independent. The notion of independence that is involved here is far from being straightforward and unproblematic. It would require a thorough analysis based on particular cases. Different types of independence, worthy of careful mapping, are likely to be involved. Now intuitively, we perceive that the independence clause covers at least two different requirements:

- On one side the requirement that the multiple derivations, as far as their *content* is concerned, be different enough (common assumptions should be as few as possible, the derivations should not be logically derivable from one another. . .).
- On the other side the requirement that, as far as their *historical origin* is concerned, the multiple derivations supporting the result *R* do not entertain a relation of mutual generation.

The independence of the derivations mentioned in the robustness scheme has, therefore, two sides:

- Independence of the historical processes corresponding to the empirical development of each derivation as an argument in favor of *R* (origin, maturation and stabilization);
- Independence of the argumentative contents of the emerging stabilized derivations.

Historical or genetic independence is rarely distinguished as such, let alone discussed, in the literature devoted to robustness. Content independence corresponds to what scholars most of the time understand and examine under the independence that is at issue in a robustness scheme or a robustness strategy. This is the kind of independence primarily discussed in Nederbragt's, Stegenga's and Trizio's chapters. Nederbragt nevertheless mentions the possibility of "the validation of one method by the other", which is a particular case of historical dependency of derivations. And his conclusion suggests a distinction similar to the one I just introduced between content and genetic independences: "absence of overlap in background knowledge of both methods and no validation of one method by the other seem to be crucial to decide on independence."

Let us further examine content independence.

1.8.1 Content (or Logico-semantic) Independence

I use the word 'content' in a very large sense: assumptions, principles, concepts, forms of argument, mathematical formula, techniques at work in each derivation. . .

At the level of contents, the discussion about independence will be based on differences that can be characterized as 'logico-semantic', or 'theoretical' in a very broad sense (e.g. not only what belongs to high level theories, but also very local, possibly implicit assumptions, etc.).

On the logico-semantic side, we have to compare the derivations with regard to symbolic resources (vocabularies, concepts, mathematical tools. . .) and the assumptions expressed with these resources. We also have to compare relations of all kinds (deduction, inclusion, analogies, reasoning schemes. . .) existing between the contents of each.

It is possible to envisage logico-semantic differences that are *more or less important:* related to more or less strong degrees of independence. Two derivations of the same result belonging to *different disciplines* (physics and biology for instance) will certainly be perceived as implying more radical differences and a stronger independence than two derivations belonging to the *same discipline* (physics for instance), and than two derivations belonging to the same specialty (particle physics for example).

At the level of the analysis of content independence, using the word 'content' in the broad sense, one could also take into account differences of (what I would call) 'epistemic spheres'. For example and typically, the fact that one derivation is seen as experimental and another is viewed as purely theoretical. Intuitively, this has an impact on the independence judgments. Intuitively and without taking into account what the derivations 'say', an experimental derivation and a purely theoretical derivation owe a part of their independence *precisely to the fact that they belong to two different spheres*. In this vein we should also examine what happens in hybrid cases such as the derivations involving simulations.

Various illustrations can be found in the different chapters of this book. In Nederbragt's chapter, in particular, the independence of the derivations involved in robustness strategies is investigated in terms of a rich panel of examples. The discussion of these examples offers many concrete instanciations of content-independence of the multiple derivations. Nederbragt analyzes content-independence in terms of differences at the level of “theoretical and methodological” “aspects”, “assumptions”, “background” and “principles” (my logico-semantic differences). We can also trace in his examples something akin to differences related to the epistemic sphere (experimental/purely theoretical investigations). Moreover, Nederbragt’s contribution points to other, finer-grained differences of this type, such as the difference between the three following kinds of study: investigation of *naturally occurring* situations (“studies of spontaneous disease cases or spontaneous exposures”); investigations of *deliberately modified natural* (in vivo) situations (“an intervention study, performed in a population in which an exposure is modified”); and investigations of *artificially-created and laboratory controlled* situations (“an experiment under controlled circumstances with animals or volunteers”). As stressed by Nederbragt, when the results of such studies converge, this increases the robustness of the convergent conclusion. I would like to add that it increases the robustness of the convergent *X*, *all other things being equal*, *specifically due to* differences concerning the *modality* of the investigation (purely observational *vs*. modifications in natural situations *vs.* experimentally created configurations). Of course, the latter dissection of the situation is for the sake of analytical clarification. In the discussion of a given scientific configuration, the different kinds of content-independence of the derivations have to be taken into account *all together* and balanced in view to an assessment of the robustness of the derived result.

1.8.2 Building an Independence Scale

As already stressed in this introduction, Section 1.4, the independence of the derivations involved in a robustness scheme comes in degrees. Moreover, intuitively, we relate degrees of independence and judgments of difference in the following way: a judgment of independence includes a judgment in terms of differences; the ‘threshold’ corresponding to independence is crossed when the derivations are *sufficiently* different; and beyond this point, higher and higher degrees of independence are involved.

With respect to this intuition, Nederbragt, in Chapter 5, provides a useful hierarchic classification of the robustness strategies according to the *kind of difference* and *degree of independence* of the derivations (see also Nederbragt (2003)). Trizio adapts Nederbragt’s taxonomy and exploits Nederbragt’s classification in his own analysis of content-independence (applied to a contemporary case study in experimental cell biology). Let me summarize this classification and reconsider it in the light of my distinction between content- and genetic-independences. In this perspective, I will categorize it as an ‘independence scale’.

Nederbragt’s hierarchy comprises four levels. The lowest, more local level refers to “confirmation of the observation when it is made for the first time and is new and surprising”. Here the experimenter’s aim is to delimit some initial conditions that will lead to stable and reproducible final conditions, and “to exclude the possibility that the observation is a coincidental artefact of the experimental manipulation of the objects of study". So the variations are minimal: the instruments and experimental configurations essentially remain the same. As a consequence, the reproducible final conditions are related to particular and local conditions. Nederbragt names this first hierarchical level "reliable process reasoning (in the experimental set-up)". Reframed in terms of my categories: at this level, we have no content-independence (the content-differences are too small); nevertheless, a certain kind of minimal historical independency is assumed, which here means that all the individual experiments involved count as truly different experiments in the sense of 'performed at different moments' (and possibly: by different experimenters).

The next level up consists in more variations of the parameters of the experiment while conserving the same kind of instrumental means. In other words, the theoretical principles involved in the conception of the instruments are not deeply modified. Nederbragt calls this "variation of independent methods", in which we have to understand "methods" as ways of playing with one and the same kind of experimental set-up: "modifications of procedure within a fixed theoretical background". At this level, the aim of the experimenter is, in Nederbragt's words, "directed at confirming the observation, to ensure that it is generalisable and that it can be made under different circumstances." Here, as at the lowest level, we still have no content-independence (although the differences in content involved are more important than at the lower level); but we do have a historical independency in the same sense as specified just above (clearly, the multiple content-similar but temporally-different individual derivations *count as more than just one of them*). Since the derivations are independent only in the sense they are conducted at different moments, and not as far as their background theories are concerned, Trizio prefers to call this second hierarchical level "variation of experimental techniques".

The third hierarchical level corresponds to "multiple derivability". Here, the theories of the experimental set-up are completely different: each derivation involves "independent methods, different in theoretical and technical background" (e.g., electron microscopy and light microscopy). Multiple derivability, as defined by Nederbragt, "is a strategy for *local* theories" (my italics, LS), for example a specific theory of bacterial invasion applied to the limited domain of particular types of cultural cells and particular bacteria. Following Trizio's exploitation of Nederbragt's scale, only at this third level do we reach a true content-independence. Clearly, a conceptual decision underlies such kinds of judgments (a decision concerning the level at which the independence threshold is situated). Indeed, nothing forbids us from saying that, at the second level of the scale, a low degree of independence is already involved.

Finally, Nederbragt's fourth level corresponds to "triangulation", which is the more general strategy, since "Multiple derivability goes from methods to theories, triangulation goes from theories to theory complexes". In triangulation, the derivations use different and independent theories, and the robust X so-produced is a theory complex.

Trizio provides diagrammatic symbols to distinguish the three bottom levels of this independence scale (as far as content-independence is concerned). He distinguishes diagrammatically the three following cases (here re-described in my terminology): *N* truly content-independent derivations (represented as *N convergent* arrows); *N* different but not content-independent derivations (represented as *N parallel* arrows); so minor variations that the result is classified, regarding its content, as one and the same derivation (representation: *one* arrow).

Although such an independence scale is highly clarifying from an analytical point of view, I would like to stress that it is not always easy, faced with a given scientific practice, to decide to which category it corresponds. The reasons are multiple. In real scientific practices, we find something like a continuum of differences rather than sharp distinctions. Correlatively, judgments of independence and 'degrees of independence' are subject to exactly the same problems than judgments of 'degrees of strength' of the derivations put forward above in Section 1.3 and generalized in Section 1.4. They are largely intuitive and qualitative judgments, in part opaque to practitioners themselves. They can vary from one individual to another one. And there is no 'objective threshold' or clear-cut demarcation between a case of independence on the one hand (i.e. a *sufficient* degree of independence, with the consequence that *two* distinct derivations should appear in our robustness scheme), and a case of dependency on the other hand (i.e., an *insufficient* degree of independence, with the consequence that the prima facie duality reduces to the unity, so that *just one* derivation should be represented in our robustness scheme). Once again, we have to face an "individuation problem". In this respect no algorithm can be applied. The analysts must make the decision based on their own judgment and experience, and here differences frequently arise.

In Chapter 9, Stegenga provides a systematic discussion of the difficulties involved in the independence problem (more precisely in the analysis of *content* independence), which clearly and extensively shows how deep these difficulties are as soon as we seek to go beyond the level of intuitions.

1.8.3 Historical (or Empirico-genetic) Independence

I now turn to the *historical* (or 'empirico-genetic') independence of the processes of emergence of the derivations.

Which kinds of scenarios can be called for in order to make sense of and to support the possibility of a genetic dependency between the convergent derivations? Several scenarios are possible, but here I will mention only one of them. This scenario seems to me, on the one hand, the best able to give some plausibility to the rather counter-intuitive idea of empirico-genetic dependency of scientific derivations, and, on the other hand, the most interesting from an epistemological point of view.

In order to grasp the principle of what is at stake, I will rely on an example carefully analyzed by Andrew Pickering in *Constructing Quarks*, one that I will here sum up in an inevitably crude and schematic way: the discovery of the weak neutral currents in the 1970s. This discovery can be explained, and is commonly explained, as the convergent result of several experiments that involve instrumental devices seen as sufficiently different, from a theoretical point of view, to be considered as content-independent. In order to capture this difference in a concise manner, I will talk about 'visual' and 'electronic' experiments.

Multiple interactions have existed between the two teams of physicists that have performed and analyzed the visual experiments on the one side and the electronic experiments on the other side. This is uncontroversial. To jump from this fact to the idea of a genetic dependence between the two experimental derivations, a supplementary thesis must be assumed: the thesis of a non-negligible "plasticity" of scientific practices.

We can understand this plasticity as a variant and an extension (as Hacking says) of the Duhem-Quine thesis. Referred to real-time practices, the thesis is *extended* in the sense that the relevant holistic units are possibly made of *heterogeneous kinds* of ingredients involved in these practices[21] – not just propositions as is the case with the Duhem-Quine thesis as traditionally presented. Plasticity is the idea that scientists looking for solidity and good symbioses can play on numerous factors and try many alignments, with the consequence that several different emergent co-stabilizations are possible – and are possibly associated with incompatible descriptions of the world.

If we are ready to concede this plasticity, it becomes plausible that the interactions between two teams involved in the resolution of similar problems with different experimental means can influence the maturation process of the derivations on each side, and therefore, at the end of the day, will also influence the stabilized couple 'derivation-result' that will eventually emerge on each side and will then be the object of an explicit description in published papers.

For example, suppose that the visual experimenters think they have succeeded in deriving the existence of weak neutral currents from their visual experiment, and that they announce the news to the electronic experimenters.[22] This will strongly encourage the latter to look for solutions that go in the same direction, to try hard to find conceptual interpretations and material modifications of their own electronic experiments that favor the convergence of the electronic conclusion with the visual conclusion.

---

[21] "Let us extend Duhem's thesis to the entire set of elements (1)–(15). Since these are different in kind, they are plastic resources (Pickering's expression) in different ways. We can (1) change questions, more commonly we modify them in mid-experiment. Data (11) can be abandoned or selected without fraud; we consider data secure when we can interpret them in the light of, among other things, systematic theory (3). (. . .) Data assessing is embarrassingly plastic. That has been long familiar to students of statistical inference in the case of data assessment and reduction, (12) and (13). (. . .) Data analysis is plastic in itself; in addition any change in topical hypotheses (4) or modelling of the apparatus (5) will lead to the introduction of new programs of data analysis" (Hacking 1992, 54). "Far from rejecting Popperian orthodoxy, we build on it, increasing our vision of things that can be 'refuted' " (Hacking 1992, 50). "The truth is that there is a play between theory and observation, but that is a miserly quarter-truth. There is a play between many things: data, theory, experiment, phenomenology, equipment, data processing" (Hacking 1992, 55).

[22] This is what happened historically. For references about this case study, see my contribution to this volume, in particular notes 1 and 6.

Of course, nothing guarantees that they will succeed (in no way does plasticity mean that experimenters can do anything they want). Nevertheless, we grasp here the possibility of a certain kind of influence that could potentially work as (what I would call a) ‘convergence inducer’.

1.8.4 Robustness, Historical Dependency, Scientific Realism and Contingentism

This possibility is interesting from an epistemological point of view, since the robustness of the convergent result *R* is not at all illusory. Practitioners have indeed found a mutual adjustment of the ingredients involved in the historical situation, which institute a stable connection between a certain experimental procedure and a certain result. But, nevertheless, the maturation processes of the two arrow-node connections involved in the emergent scheme of robustness, namely ‘visual derivation-*R*’ on the one side and ‘electronic derivation-*R*’ on the other side, cannot be said to be independent as far as their historical genesis is concerned.

From an epistemological point of view, this has implications for the realist/constructivist issue. The robustness scheme is currently described as a *criterion* for realist attributions, objectivity and the like. In this vein Wimsatt writes: “Robustness is widely used as a criterion for the reality or trustworthiness of the thing which is said to be robust” (Chapter 2, p. 74). “Robustness is a criterion for the reality of entities” (Chapter 2, p. 76), “a criterion for objecthood” (Chapter 2, p. 75). As I develop in Chapter 10, Section 10.3, the (often implicit) argument that lurks behind the inference from robustness to truth or reality attributions is one version or another of the well-known “no-miracle argument” commonly invoked by realists. Or, in Stegenga’s words:

> one way to understand robustness is as a no-miracles argument: it would be a miracle if concordant multimodal evidence supported a hypothesis and the hypothesis were not true; we do not accept miracles as compelling explanations; thus, when concordant multimodal evidence supports a hypothesis, we have strong grounds to believe that it is true.

As a description of the frequent ‘tacit inferences’ followed by real scientists, I take Wimsatt’s quotations above, and similar claims, to be undeniable. But *as a philosophical thesis about the realist pretention of our science* and the fact that robustness is truly a strong argument in favor of the realism of scientific results, we have to be more cautious.

Indeed, if, historically, the multiple derivations and their convergent result can be seen as co-adjusted and co-stabilized (in senses that should be specified case by case), the miracle of the convergence appears definitely less miraculous. Or rather, it appears to be a miracle of a different kind from the one involved in the traditional so-called “no-miracle argument” of the realist (see Chapter 10, Section 10.3 and the last part of Section 10.16). The miracle would rather be that practitioners have succeeded – under the various constraints, perceived as extremely strong, which were present in a given historical situation – to obtain not only one stabilization, but, moreover, a stabilization that converges with other, already available ones.

I think we should continue to speak of a miracle, in order to point to the extreme difficulty of what is at stake and to the effort and shrewdness it might require. But, obviously, this type of miracle no longer feeds realist intuitions. We are less inclined, not to say no longer inclined, to call for the external pressure of reality as the origin, if not the cause, of the convergence – even if one might choose to conserve the idea of reality as a global source of non-isolable constraints and as a regulative ideal.

What is weakened, in this scenario, is the almost irresistible explanation of robustness in terms of realism: the equation 'it's robust, therefore it is true' (in the sense of correspondence truth). What is moreover weakened – although I cannot really argue this point here – is the equation 'it's robust, therefore it was inevitable' (given certain conditions of possibility working as initial conditions).

So the philosophical questions lurking behind all of this are the realism *versus* antirealism issue and the "inevitabilism" *versus* "contingentism" issue (in Hacking's terminology[23]). Several chapters of the present book approach these questions.

Regarding the contingentism/inevitabilism antagonism, Pickering is one of the analysts of scientific practices who has gone the furthest in the discussion of the contingency thesis. In previous works (see especially Pickering (1995, 201X)), he has articulated the idea of a genuine contingency of scientific achievements and advocated it in relation to various historical configurations. In his contribution to the present volume, he reaffirms the contingency of scientific achievements, be they experimental facts, material devices or more abstract conceptual systems, at the very same time that he vindicates their very robustness. On the basis of his account of scientific development in terms of symbioses, what could forbid several very different and conceptually irreconcilable global fits, all of them achieving a good machinic grip on reality? And why would this multiplicity be a problem? This is how I understand the moral of Pickering's chapter: robust but not unique. The history of science is a contingent process. The results achieved along the way depend on this process, since they acquire the status of results on the basis of their inter-stabilization with other co-present ingredients, on the basis of "contingent and emergent productive alignments". They cannot be detached from this process: they are "path-dependent". Hence, scientific results, although robust, both epistemologically (in the sense that a satisfying symbiosis has indeed been achieved) and ontologically (in the sense that the world or a true "otherness" indeed enter in them constitutively), are nevertheless contingent (in the sense that they could have been different and incompatible).

Nickles and I, both inspired by Pickering's work, will also approach the contingency issue in our contributions.

---

[23] See Hacking (1999, 2000). For further consideration of the inevitabilitist/contingentist issue and relevant references see Soler (2008b) (for the presentation of the problem and its situation in the landscape of the Science Studies) and Soler (2008c) (for an analysis of the structure of the problem and its internal difficulties). See also Soler (201X) for a general argument in favor of contingentism. In french, see Soler (2006a).

As for the realist issue, and especially the passage from robustness to truth, many chapters will discuss it: the three just-mentioned ones of Pickering, Nickles, and my own, where the realist issue is mobilized in close relation to the contingentist issue; Stegenga's chapter, as already noted and quoted at the beginning of this section; and Boon's article, in which the author argues that the jump from robustness to the truth of theories and the reality of scientific entities is not legitimate.

1.9 Sequential Overview of the Contents of This Book

1.9.1 Chapters 2 and 3: Wimsatt on Robustness, Past and Present

As already noted at the beginning of this introduction, William Wimsatt is one of the pioneers of the topic of robustness in general and the person most important for making it a topic for philosophy of science. His seminal 1981 paper, "**Robustness, Reliability, and Overdetermination**", is reprinted in this volume as Chapter 2. Since then he has added material in other papers and in his 2007 book, *Re-Engineering Philosophy for Limited Beings*. Wimsatt deserves credit not only for introducing the topic but also for providing an analysis of, and a schema for, robustness that has been, along with Richard Levins' 1966 paper, "The Strategy of Model Building in Population Biology", the main focus of ensuing discussions of robustness. The contributors to this volume maintain that focus.

In his new chapter for this volume, **"Robustness: Material, and Inferential, in the Natural and Human Sciences", Wimsatt** recounts the origins of his concern with robustness, as he experienced it at the beginning of the 1970s, in the context of attempts to connect more closely the domains of population genetics and community ecology. The fact that domains so closely related in topic were employing quite different kinds of models was puzzling. Wimsatt describes the complex theoretical situation at the time and the inadequacy of the methods developed by philosophers of science with respect to the biologists' aims. All this pushed practitioners of the field to elaborate strategies related to robustness. Among others, it motivated Richard Levins to develop the idea of robust theorems that are stable across multiple models, and it led Donald Campbell to introduce the related notion of "triangulation". Inspired by these different approaches, Wimsatt formulated a unified and systematized analysis of robustness. This permitted him to identify various kinds of failures of robustness, such as the "pseudo-robustness" that results when the independence of convergent approaches first assumed proves to be illusory.

Wimsatt's new contribution for this volume provides useful tools toward the recognition and differentiation of types of robustness. In particular, he distinguishes two main types of robustness, inferential and material.

- In inferential robustness the target is what I would call a symbolic object, typically a scientific proposition. In this case, establishing robustness consists in designing multiple inferential means to the target proposition. Wimsatt proceeds to distinguish three subcases, according to the kind of inferential means involved:

(1) empirical means available to the empirical sciences ("sensory modalities", "instrument mediated" derivations, etc.); (2) empirical tools employed by the social and "human intentional sciences" "that make intensive use of intentional responses" such as questionnaires; and (3) analytical means consisting of "multiple independently motivated and constructed models", mathematical derivations, and the like.

- Material robustness concerns robustness attributions to material systems – typically biological organisms designed by nature, but possibly including technological systems or non-living objects and their behaviors – that acquire their robustness through natural processes. Wimsatt's chapter mentions other classification attempts, some very recent, and situates them with respect to his proposal.

In addition to these analytical tools, Wimsatt analyzes a case of material robustness through a puzzling example. He calls it "the paradox of sex" and takes it to be "the most interesting (and indeed most focal) problem involving robustness in biological organisms". The problem involves two levels: on the one hand the inferior genetic level and the important variation it manifests (genetic variability, genetic recombinations and the like), and on the other hand the superior level of phenotypic properties. The enigma is the following: how is the stability at the superior level possible (leading to members of the species that are sufficiently similar to be able to have sex together), given the high variability of the inferior level? In response to this puzzle, Wimsatt discusses what he calls a "sloppy gappy robustness" or "statistical robustness" in which the desired result at the higher level must only occur sufficiently frequently, although not in all cases.

1.9.2 Chapters 4, 5, 6, 7 and 8: Case Studies of the Robustness of a Single Node

The book continues with five chapters centered on (more or less simple and more or less prototypical) case studies devoted to the robustness of a relatively simple node *X* (a scientific proposition or a scientific image), both in the domain of the empirical sciences (Chapters 4, 5, 6 and 7) and in mathematics (Chapter 8).

A. In Chapter 4, **"Achieving Robustness to Confirm Controversial Hypotheses: A Case Study in Cell Biology", Emiliano Trizio** takes up a relatively simple case, although not completely prototypical of the Wimsattian robustness scheme, and a little bit more complicated than this scheme. The analysis of the case by Trizio enables us, both to meet the shift between real practices and idealized schemes, and, in feedback, to understand better the prototypical simple scheme. Moreover, it provides a clear illustration of the degrees of difference that might be involved between the multiple derivations constitutive of a robustness scheme.

   The case concerns bacterial endocytosis in mammalian cells. Endocytosis is a process by which cells engulf and absorb external material such as proteins and bacteria. The question at issue is how large these external "particles" can be. In 2005, some scientists claimed to have shown experimentally, by fluorescence microscopy, that endocytosis can involve particles of a very large size. This claim was in tension with the theoretical beliefs at the time. Many scientists reacted with skepticism to it. They asked for further experiments that use different techniques from fluorescence microscopy. This amounts to a demand for robustness, in a situation where the experimentally derived results challenged previous theoretical postulates about the size of the particles involved. The demand for robustness is a demand that the results of both practices of fluorescence microscopy and electronic microscopy be brought in line with one another, that is, that they converge toward a common result.

The case study presents several interesting features in addition to the above-mentioned ones. For one thing, it is an investigation of ongoing research that is still open. Trizio's chapter analyzes a contemporary scientific practice in real-time, thus avoiding the risk of a posteriori reconstructions of robustness schemes that ignore the uncertainties or forget the messiness of discordant evidence. It is moreover a fascinating example of broadly wimsattian robustness worries manifesting themselves in the heated action of scientific research. The demand of robustness appears as an explicit criterion for the acceptance of experimental results: the referees of the journal *Nature* ask for it to be satisfied and reject the paper that lacks it.

A curious feature of this case is that the old understanding of the limit on size of particles that can be absorbed by means of endocytosis had gained a kind of theoretical status, informing much other research, despite its not being part of a well-confirmed theoretical mechanism or model in biochemistry. Rather, its provenance is related to technical contingencies: previously used detection techniques never turned up cases of large particles being absorbed. Some of the scientists involved in the debate therefore regard it as a "dogma" produced by a contingent prior history of observation based on older detection methods.[24]

After laying out the endocytosis controversy, Trizio applies to it an adapted version of Nederbragt's classification of the experimental derivations according to the kind of difference and degree of independence they manifest (see above Section 1.8.1). As a result, three kinds of experimental derivations appear to be involved in the historical situation. Some of the experimental actions are categorized as variations in the parameters of the experiment without essentially altering it. Other experimental actions are analyzed as cases of significant differences in experimental techniques but without altering the underlying theoretical principles. In such a case we are intuitively inclined to say that different but not independent derivations are involved. Finally, a third configuration is also found, which corresponds to Wimsatt's definition of robustness: experimental

[24] If they are right we can suspect (although this is not articulated in Chapter 4) that there is a kind of historico-social entrenchment in operation here. Wimsatt's own analysis suggests that entrenchment possesses a historical dimension. On social entrenchment, see Chapter 5.

actions for which the theoretical principles underlying the experiment and its techniques are "entirely different" (for example fluorescence microscopy and electron microscopy).[25] In that case we are inclined to say that genuinely independent derivations are in presence. Trizio provides diagrammatic symbols to distinguish these three cases, and using them, he proposes a schematic arrows-nodes representation of the situation which immediately and very clearly shows in what respects this case is more complex than the Wimsattian configuration.

Although Trizio bases his analysis on a particular case, it would seem to generalize quite easily. As he says, "This complexity is likely to be widespread across different scientific disciplines, and the neat, idealized convergence described by the classical robustness scheme should be an exception rather than the rule (. . .)". Most often, the wimsattian scheme is useful as a regulative ideal or (in Trizio's suggestive phrase) as a "methodological attractor" for practitioners.

B. Like Trizio, **Hubertus Nederbragt**, "**Multiple Derivability and the Reliability and Stabilization of Theories**," discusses the problem of robustness in the biomedical sciences. His central case study is the invasion of cultured cells by bacteria, but he employs many other examples.

Nederbragt begins by distinguishing three logical situations that are often lumped together: consilience of inductions, multiple derivability, and triangulation. He argues that these serve different purposes in scientific research. He spells out the differences he sees, then focuses on multiple derivability and triangulation in his discussion of cases.

Multiple derivability and triangulation are two different robustness strategies. More precisely, both are *inductive* strategies for the purpose of theory improvement. Both strategies make the theory more reliable when they succeed. The difference lies: (a) in the degree of generality of the targeted theory at stake: for multiple determination, local theories (for example a specific theory of bacterial invasion applied to the limited domain of particular types of cultural cells); for triangulation, "interdisciplinary theories or theory complexes", which "may even lead to new disciplines". (b) In the nature of the independent derivations involved: theoretically and technically independent methods for multiple determination; theories for triangulation.

With this categorization, Nederbragt adopts a narrower, more restricted sense of "multi-determination" than the one involved in Wimsatt's terminology (or in my own taxonomy as presented above in this introduction – not surprisingly since my taxonomy has been built from and in continuity with Wimsatt's). In Wimsatt's framework as well as in my conceptualization, Nederbragt's triangulation would be described as a case of multi-determination. Both Nederbragt's multi-determination and Nederbragt's triangulation would instantiate a robustness scheme defined as 'convergence under multiple determinations'. Indeed, in both cases we have several taken-as-solid derivations-arrows converging on one and the same result R taken-as-robust in virtue of this structural configuration.

[25] For problems of individuation that arise here, see Chapter 9.

The difference lies in the ingredients of the scheme (and hence in the aim of the strategy if we consider the situation dynamically): *R* refers either to a local theory or a theory complex; and the arrow-derivations refer either to material-experimental methods or to intellectual-theoretical derivations.

Nederbragt's chapter continues with an analysis of the independence of the derivations involved in robustness strategies, and develops in this context a hierarchical classification of different types of derivations and robustness strategies (summarized and commented above Section 1.8.1). "A hierarchy of various types of derivations may be discerned in scientific practice", Nederbragt writes, from very particular ones to more and more general ones, which, respectively, lead to more and more robust derived theories. Briefly put: "a hierarchy of derivabilities" is conceptualized into a "hierarchical model of theory making".

Finally, Nederbragt's chapter discusses what might happen in historical situations in which robustness strategies like multiple determination and triangulation fail to work, either because no more than one derivation is available, or because the multiple available derivations are too weak to achieve genuine robustness. In such situations there are schematically two possibilities according to Nederbragt. In the first, the *X* that was a candidate to robustness is finally abandoned or rejected as fragile. In the second, the *X* that was a candidate for robustness is stabilized through "social anchoring", that is, "because of social interaction between method, theory and scientists of the research group". The contrast between the two possibilities leads Nederbragt to introduce a distinction between reliability and stability. An element *X* of scientific practices can be historically stabilized because of different kinds of factors: different "ways of anchoring of knowledge may be distinguished, leading to its stabilization". Reliability (and robustness is a case in point) is stabilization "by epistemological arguments and factors" or "epistemological anchoring". But stabilization can also occurs "by social, i.e. non-epistemological, factors". As a consequence, stability is a more general category than reliability. According to Nederbragt, all kinds of combinations can be found among actual historical cases: "Theories may be reliable but not stable because they do not fit in a reigning paradigm. (...) Theories may be stable, although they are epistemologically unreliable."

C. In **"Robustness of an Experimental Result: The Example of the Tests of Bell's Inequalities", Catherine Dufour** case study concerns experimental tests of the inequalities following from Bell's theorem of 1964. This theorem states that if standard quantum mechanics (SQM) is correct, then quantum entanglement experiments of the kind first proposed by Einstein, Podolsky, and Rosen in 1935 should show that physical nature violates the inequalities in specified circumstances; whereas if local hidden-variable theories (LHVT) is correct, then nature should satisfy the inequalities. So the experimental tests of Bell's inequalities amount to tests of SQM over LHVT (or Bohr against Einstein as it is often said).

Dufour argues that even the apparent convergence of the so many experimental outcomes supporting SQM over LHVT is not quite enough to meet the wimsattian robustness ideal.

Over the past several decades, physicists have performed a long series of such tests of gradually increasing sophistication. However, two notable loopholes remain: the so-called “locality loophole” and the “detection loophole”. The locality loophole can be closed only by guaranteeing that the two separating ‘particles’ have no means of ‘communicating’ with one another, even by a speed-of-light signal. The detection loophole is a specific version of a general problem that affects many kinds of experiments in the sciences. In a typical test of Bell’s inequalities, only a sampling of the particles (usually ranging from 20 to 80%) are actually detected.

While physicists and technicians have made significant progress in closing both loopholes, it is exceedingly difficult, perhaps even impossible, to close both at once. Dufour emphasizes that even the most recent sophisticated experimental designs still fall short of the EPR ideal. Many, perhaps most, physicists who care about this issue believe that the tests decisively support SQM, but doubters remain able to contrive LHVT arrangements that evade this conclusion. Leading physicists themselves disagree about how much robustness is enough (that is, about the desirability or the necessity of supplementary experimental derivations). Accordingly, the robustness debate remains somewhat open in this case. In other words, it would be premature, according to Dufour, to conclude that LHVT has been empirically refuted, because questions of robustness remain unresolved. An element of fragility remains.

The analysis of the case shows how judgments about the robustness of an experimental *result* are dependent on and intertwined with judgments about the solidity of the *derivations*.

In such a case we can ask why, despite the concordance of the so many experimental outcomes obtained, physicists did not stop and a so high number of experiments have been performed. As noted by Dufour, part of the answer lies in the importance of the theoretical implications for our scientific worldview. Beyond the particular case, the general moral is, I think, that the number of derivations deemed sufficient to provide enough robustness to a result on which they converge depends on contextual factors, ‘external’ to the robustness scheme itself (here the circumstance that the experimental result is perceived as especially important).

Another interesting feature of the Bell test experiments is that one experimental derivation *against* SQM was developed but never published. Even its authors (Holt and Pipkin) rejected its conclusion. This result has been set aside almost as if it never existed. The methodological-philosophical question of how to handle ‘outliers’ is a pervasive problem in the sciences. Given the universal significance of the debate between SQM and LHVT, the problem is all the more interesting in this case.

Finally, the historical situation under scrutiny exemplifies a non-prototypical robustness scheme, non-prototypical but interesting because it is so frequently instantiated in the history of science. It is a situation in which the multiple experimental derivations involved in the temporal sequence are thought as several classes of successive generations of improved experiments of one and a same kind. Although beyond the scope of Dufour's chapter, she provides material toward a future, finer-grained reflection on the kinds of independence required in a solidity scheme. Clearly, successive experiments thought as improvements of one and a same kind of experiment are not independent in the same sense as multiple experiments of different kinds. Nevertheless, there is a sense in which the experiments involved in the first configuration are considered as independent (this sense is related to the "empirico-genetic independence" introduced above Section 1.8). One manifestation of this independence is that an improved experiment does not completely 'cancel' the anterior ones, as if just the last one counted. Although the new and old experiments are clearly not on the same footing, although the new-and-improved one is clearly more important and compelling than the preceding ones, the fact that *two* (or more) such experiments led to the same conclusion counts for more than does the last experiment in the series, standing alone. There is a gain provided by the very fact of a convergent multiplicity. In a sense the more sophisticated (and often more precise) experiments later in the series confirm the promise indicated by the earlier, cruder efforts. Scientists get the sense that the entire series is thus 'on the right track.'

D. In Chapter 7, **Catherine Allamel-Raffin and Jean-Luc Gangloff**, informed by the ethnographic work of the first author at the Harvard-Smithsonian Center for Astrophysics, analyze in detail the argumentative structure of a single, recent (2001), frequently cited radio-astronomy paper. The scientific paper presented two large maps of the CO distribution in the Milky Way Galaxy and the argumentative justification of their accuracy.

As the title, "**Scientific Images and Robustness**" of their contribution suggests, Allamel-Raffin and Gangloff argue that images and other nonverbal features of scientific articles play a more than illustrative role. Their second major claim is that the procedures employed by the radio astronomers to convince themselves and their readers that the Milky Way maps are accurate satisfy Wimsatt's strategy for achieving robustness.

The originality of their case study consists in showing that nonverbal entities – Images, maps, diagrams, etc. – can be the objects of robustness claims, that is, the conclusions of converging series of arguments, each constituting a derivation in the sense defined in this introduction. Moreover, these derivations sometimes themselves employ images as well.

It is noteworthy that the scientific paper in question is only ten pages long yet contains twenty images of various kinds. Allamel-Raffin and Gangloff emphasize the importance of images in the economy of scientific argumentation. A map containing many points, for instance, is worth more than a thousand words, for the information it contains could only be expressed in terms of an incredibly large number of sentences, in which form it would hardly be intelligible.

All this led the authors to see Wimsattian robustness analysis as a bridge between philosophy and the sciences. They conclude that

> (. . .) the concept of robustness analysis (which is in fact a philosophical concept) gives a perfect account of the procedures used in the day-to-day activities of a lab to prove scientific assertions. (. . .) With the robustness concept, we have a perfect example of a 'working' concept that can build a bridge between the scientists and the philosophers.

E. Finally, in Chapter 8, **"Are we still Babylonians? The structure of the foundations of mathematics from a Wimsattian perspective," Ralf Krömer** discusses an important but rarely investigated issue, namely, whether and how robustness is involved in mathematical practice. Does robustness analysis have anything to contribute to this domain?

A preliminary indication that it does is the fact that several important mathematical results have been proved in multiple ways.[26] To be sure, in some cases the importance of a mathematical proof lies in the connection it forges between two different domains. But in some of these same, as well as other, cases, additional proofs apparently add solidity to the theorem asserted and, indirectly, to the methods by which it was previously proved.

Yet, as Krömer indicates, according to "the mainstream view of the epistemology of mathematical proof", the answer to our opening question seems to be a clear 'no'. This is because a mathematical proof, typically conceived as a deductive proof from a set of explicit axioms and previously proved theorems, is taken to be absolutely certain and inescapable – quite in contrast to inductive inferences in the empirical sciences. If you possess such a proof, the resulting theorem is necessarily established, once and for all, irreversibly. Thus no additional support is needed: the maximum is already achieved. Adding anything else would be wasted effort.

If this conception of mathematics does exhaust the practice of mathematical demonstration, then a robustness strategy, defined as developing multiple independent mathematical derivations of one and the same result, gains no foothold in mathematics. Robustness would neither describe a desired aim of real mathematical practitioners, nor would it be a fruitful site for the epistemological analysis of mathematical practice. Thus we could point to an essential difference between demonstrational practices in the formal sciences on the one hand and in the empirical sciences on the other hand. And, as Krömer notes at the beginning of his chapter, that is precisely the traditional conception of these two enterprises.

Krömer's contribution clearly shows that this common view, at the very least, does not exhaust the situation, which turns out to be more complicated and interesting than anticipated. He advances several challenges to the mainstream view. First, as soon as one examines actual practices rather than traditional idealizations, it appears, as a matter of fact, that mathematicians do indeed strive to re-prove theorems. And when the analyst of science asks why they do so – as

---

[26] More generally, students of history should be curious why, even in wider contexts, such as 'proofs' of the existence of God, thinkers thought it necessary to provide more than one proof, each one supposedly decisive in itself.

John Dawson did independently of any connection to the robustness issue – one important part of the answer involves a robustness scheme essentially similar to the one valued by practitioners of the empirical sciences. In fact, the idea of an absolute necessity attached to mathematical proofs has sometimes appeared to be an illusion. As the history of mathematics shows, "Proofs can be false, and errors can pass undetected for a long time". Thus mathematicians historically and today experience *degrees* of confidence in mathematical derivations, hence their search for different derivations that, taken together, would enhance the confidence in a given mathematical result. As a consequence – and as already suggested by Wimsatt (1981), himself inspired by Richard Feynman – the edifice of mathematics is less "Euclidean" and more "Babylonian" than is commonly thought. In other words, there exist over-connected mathematical structures: structures in which many propositions are at the centre of multiple derivations; in which multiple, non-hierarchical paths are possible from one proposition to the other; in which no special set of propositions is identified as *the* ground of the whole edifice.

From this opening line of thought, Krömer draws a first – in my opinion very important – conclusion with respect to the desirable re-orientation of the philosophy of mathematics: "we might (and should) be led to feel the need for a more subtle, and more appropriate, epistemological conception of mathematical proof and its role for conviction (eventually making use of the concept of robustness)."[27] Just as many philosophers now agree that good philosophy of the empirical sciences requires close attention to actual scientific practices, so, philosophers of mathematics need to pay more attention to the actual practices manifested in mathematical work-in-progress.

Second, there are interesting mathematical propositions for which mathematicians do not possess a formal deductive proof and for which they doubt that they will ever have one. Typically, such claims are either taken to be independent of a preferred axiomatic base or else (in some cases) being considered as an addition to an existing base, whence worries about the consistency of the system arise. Krömer considers two mathematical propositions of this kind, one in the well-known context of century-old set theory, the other in the context of a less well-known and much younger mathematical framework, namely category theory.

What processes could secure, or at least enhance, the reliability of such propositions in the absence of deductive proofs? Schemes akin to robustness are, or could be, involved in the answer. In the case of set theory, Krömer considers an answer proposed by Bourbaki: the proposition can be taken as robust in the absence of a deductive proof since it has been applied so many times in so many branches of mathematics without producing any contradiction. Krömer discusses

[27] In this respect, see Van Bendegem et al. (2012). Krömer's conclusion just above exactly expresses the spirit in which this book, entitled *From Practice to Results in Logic and Mathematics*, has been impulsed, following a Conference organized by the PratiScienS group and myself in June 2010. See http://poincare.univ-nancy2.fr/PratiScienS/Activites/?contentId=6987.

the similarities and differences between such a proposal and the Wimsattian robustness scheme as the latter applies to the empirical sciences. He also points to difficulties, some of which are specific to the mathematical field, and some shared with empirical science (for example, the hard problem of the independence of multiple derivations). In the case of category theory, where a solution à la Bourbaki is unavailable because the theory is still in the early stages of its development, Krömer furnishes reasons why the search for multiple derivability could be a valuable strategy for mathematicians. This is an especially important suggestion, given the present situation in which category theory presents us with foundational problems yet is employed as the only available demonstration of some important propositions.

1.9.3 Chapter 9: A Systematic Panoramic Analysis of the Robustness Notion

In Chapter 9 **"*Rerum Concordia Discors:* Robustness and Discordant Multimodal Evidence", Jacob Stegenga** offers a systematic and rigorous panoramic analysis of the robustness idea and the main difficulties associated with evaluating robustness, and provides multiple illustrations relevant to diverse empirical sciences. Hopefully, such an analysis will be better understood, it's relevance and usefulness will appear more clearly, once having in mind the multiple concrete cases considered in the previous chapters.

After having investigated some uses of "robustness" in the philosophy of science literature, Stegenga builds his own general definition and terminology: "A hypothesis is robust if and only if it is supported by concordant multimodal evidence." In this definition, "hypothesis" is used in a broad, generic sense, intended to cover all kinds of scientific knowledge components about which robustness can be predicated: experimental results, statistical analyses, models or any kind of scientific method. In the same spirit, "multimodal evidence" aims to encompass a great diversity of "way[s] of finding out about the world": human sensory modalities, experiments, statistical analyses, comparative observations, mathematical models, heterogeneous techniques and so on.

Beyond the general definition, "robustness-style arguments" will possibly differ in many respects: with respect to the kind of modes and the kind of hypotheses they involve; the number of different modes involved; the strength of the evidence that each provides; the kind and degree of independence that the multiple modes, considered together, show; the quality of the concordance they manifest; and how to handle the existence of discordant evidence. With the elucidation of these different variables, Stegenga greatly clarifies what robustness-style arguments are made of, and exhibits by the way all the complexity of the evaluation involved. Moreover, he distinguishes robustness from "Another way in which multimodal evidence is said to be valuable", called "security", which "is the use of one mode of evidence to support an auxiliary hypothesis for another mode of evidence, which is itself evidence for the main hypothesis of interest".

Stegenga points to the hopes invested in robustness-style arguments and lists the numerous valued epistemic tasks that they have been credited to accomplish (see above Section 1.1). But immediately after, he stresses that these hopes and tasks are in need of philosophical analysis and justification. And it is clear from Stegenga's incisive critical examination that many of them raise questions, to say the least. Stegenga articulates several difficulties that I take to be deep and often ignored or minimized.

Although it is indeed a "hard problem", I pass rapidly over what Stegenga presents, rightly I think, as two empirical facts, namely, that multimodal concordant evidence is rare and that most of the time multimodal evidence is discordant, producing various sorts of incongruity or outright inconsistency. Rather, I shall concentrate on the analytical difficulties.

First, when we attempt to get beyond vague intuitions, it is not at all easy to specify what multimodal evidence is. What defines an individual mode as *one* mode and distinguishes it as sufficiently independent from another one? Stegenga calls this difficulty "the *individuation problem* for multimodal evidence". We could call it as well the independence problem. As already noted Section 1.3, this problem is crucial since its solution determines the *number* of different modes (or derivations-arrows) involved in the robustness scheme associated with a given historical situation, and since the evaluation of this multiplicity is the first logical step of robustness assessments. (I speak here of a 'logical step', not 'temporal step', because I think that, in practice, the different difficulties analytically distinguished by Stegenga are not treated one after the other but mixed all together). This is the problem of describing the nature of independence, as well as of justifying, on the one hand the degree of independence the modes should possess in order to count as providers of robustness (total? partial?), and on the other hand the criteria according to which the independence between modes can be assessed. To belong to independent modes, is it sufficient that the arguments differ only in a single background assumption, while sharing the rest? At the other extreme, must they have no background assumptions in common? Or is it enough that they differ in at least one problematic assumption? These are the kinds of questions that must be addressed. In addition to the *choice* of the criteria, their *application* to concrete cases presupposes that we are able to *recognize* all the background assumptions involved in the configuration under scrutiny, which is a strong – and I would say, in agreement with Stegenga, not very realistic – presupposition (just think of the related problem, to which Pierre Duhem called attention, of identifying all the auxiliary assumptions involved in a typical experimental test).

I see the independence problem as a part of a second, more general problem put forward by Stegenga, "the amalgamation problem" ("*how* multimodal evidence should be assessed and combined to provide systematic constraint on our belief in a hypothesis"). This is a balance problem, the solution to which would determine a global judgment about the robustness of the hypothesis under discussion. Stegenga sees the task as the determination of an "amalgamation function" and sketches what its task would be:

> [A]n amalgamation function for multimodal evidence should do the following: evidence from multiple modes should be assessed on prior criteria (quality of mode), relative criteria (relevance of mode to a given hypothesis) and posterior criteria (salience of evidence from particular modes and concordance/discordance of evidence between modes); the assessed evidence should be amalgamated; and the output of the function should be a constraint on our justified credence.

One important problem facing the task of constructing such an amalgamation function is that, as a matter of fact, practitioners often disagree about these evaluations.

> The basis of many scientific controversies can be construed as disputes about differential assessments of these desiderata: one group of scientists might believe that evidence from some techniques is of higher quality or is more relevant to the hypothesis or has greater confirmational salience than other techniques, while another group of scientists might believe that evidence from the latter techniques is of higher quality or is more relevant or salient.

Stegenga seems to suggest that this kind of problem might be overcome, when he concludes: "The construction and evaluation of such schemes should be a major task for theoretical scientists and philosophers of science." I am not so confident and suspect, rather, that there is no such general and systematic function. I cannot see how we could get rid of the intuitive, judgmental, non-computational character of global estimations of scientists and the historical fact of individually varying appreciations regarding these matters. Although Stegenga's analysis is immensely helpful with respect to our conceptualization of the robustness problem, and although such general schemes as the Wimsatt robustness scheme are analytically clarifying, I am afraid they will not give us anything like a general, uniform decision algorithm. In my view, individually-variable intuitions and assessments are an ineliminable part of real scientific practices.

1.9.4 Chapters 10 and 11: The Solidity of Derivations

A. **Léna Soler's** chapter,[28] "**Robustness of Results and Robustness of Derivations: the Internal Architecture of a Solid Experimental Proof**", analyzes in detail an influential scientific paper usually considered as one of those which have contributed to the discovery of weak neutral currents (one of the developments discussed by Pickering in *Constructing Quarks*). The chapter is based on work done at Gargamelle, the giant bubble chamber at CERN. This work was important, since the discovery of weak neutral currents (weak NCs) was consistent with the so-called Standard Model of Glashow, Salam, and Weinberg, the particle theory that unites the electromagnetic with the weak and strong nuclear forces. Soler's focus is on derivations more than final results, that is, on the "argumentative line" that leads to the result. The Gargamelle

[28] This chapter is here presented by Thomas Nickles.

work was one of three different experimental argumentative lines that converged to robustly constitute the 'discovery' of weak neutral currents in the early 1970s, the other two coming from Aachen and Fermilab National Accelerator Laboratory. Each of these argumentative lines is represented by a single arrow in Wimsatt's scheme, with the arrows converging on the same result: neutral currents exist.

Soler's strategy is to take Wimsatt's robustness "panoramic" scheme as her starting point and then to zoom in on the Gargamelle arrow, with the intention to look at the detailed practices necessary to produce it. This investigation shows that, while Wimsatt's scheme remains useful as an idealization, appreciation for the detailed scientific practices involved in realizing it (themselves frequently employing that scheme in micro-contexts) somewhat undercuts the common claim that robustness considerations diminish the contingency of scientific work to the point that we can regard mature scientific results as inevitable. In short, there remains a leap from 'robust' to 'true'.

Soler shows that, as we zoom in on the Gargamelle arrow, we find that this simple representation conceals a multitude of black boxes or modules involving solidity/robustness arguments, some inside the others like a set of Russian dolls (although with interactions among them). Her purpose is to open these black boxes. The resulting story (which she insists is still greatly simplified) becomes a dizzying spiral of complex stories within stories, reaching back into scientific history, and shows how naïve are familiar philosophical accounts of scientific experimentation and the corresponding formal confirmation theories. Although introducing levels of complexity hidden from most philosophers in their accounts (and from some of the scientists themselves), Soler points out that her analysis is still located at "a level of scientific practices that is emergent with respect to laboratory practices themselves."

The Gargamelle experiment involved making and interpreting photographs of particle interactions inside the large bubble chamber. The general point to be made here is that the sought-for weak neutral current reaction, ν + *nucleon* $\rightarrow$ ν + *hadrons*, where ν is a neutrino, was very far from something 'given' by observing nature. Soler outlines the constructive steps and expert judgments required at every step of the way. The raw data consisted of 290,000 photographs of events of possible interest. But how was this data to be analyzed and interpreted? One problem is that neutral particles leave no tracks in bubble chambers or on film, and neutrinos are neutral. Another (related) problem, the neutron-background problem, was to distinguish pseudo-neutral-current events from those produced by high-energy neutrons (also neutral particles). Handling this difficulty required a problem shift: to determine the *ratio* of NC events to charged-current events. And this is just the beginning of the story. To make sense of the complexities, Soler provides an architectural metaphor with accompanying figures: four floors of data analysis.

The story of the work done on each of these 'floors' is, again, quite complicated, with robustness considerations often central. For example, the ground floor involved four kinds of data filtering, one of them being an energy "cut" at 1 GeV to weed out many pseudo-events among the tracks too ambiguous to decipher.

> The filtering operations aim at eliminating some confusions, but they can themselves be sources of mistakes. For example, if the energy cut at 1 GeV is too severe, real NC-events might be artificially eliminated, and the risk is to conclude mistakenly that weak neutral currents do not exist. But if the cut is too permissive, too many pseudos might be taken for authentic NC-events, and the risk is, this time, to conclude mistakenly that weak neutral currents do exist.

So the attempt to make the argumentative line solid can actually introduce new elements of fragility (a theme that Nickles takes up in a different context). A kind of compromise or equilibrium is sought that makes the result solid *enough* to proceed.

On the third floor, the way scientists dealt with another problem, the muon noise problem, looks at first like it fits the Wimsatt model of robustness, with three lines of argument converging on the final conclusion that then became input data to the next level of analysis. But, as is typical in actual research, the three results were not identical. Finding sufficient concordance required further work (which in this case is relatively invisible but nevertheless present), work involving what Soler calls "calibrating re-descriptions" and inferential moves, based on scientific judgment, that were not logically obligatory. Here Soler's chapter intersects Stegenga's and others. Again at this level of description, the concordance of results is not simply "given." Rather, it must be constructed. There are severe constraints on these constructions, to be sure, but there is also room for flexibility.

Soler concludes that the solidity of a derivation consists of two components. One is the internal solidity of that argumentative line, based on the remarkable amount of detailed work from which it emerges. The other is its relation to external or extrinsic circumstances such as the existence of other lines of argument. Here the solidity comes from something like Wimsatt's elementary robustness scheme rather than from the internal detail of a single argumentative line.

In the final sections of her chapter, Soler points to the implications of her work. One, already mentioned, is the role of expert judgment. She clearly believes that formal algorithms are insufficient to model scientific practice. "To account for such judgments (as far as this can be done), the philosopher will have to take into account the particular content to which the robustness skeleton is associated in each case." This point applies also to any attempt to include an algorithmic version of Wimsatt's scheme in a formal confirmation theory. There is no getting around the need for expert judgment – human decision, human agency – at many stages of research (again, an intersection with Pickering's chapter).

The most important implication, in her eyes, is the probable historical path-dependence of science and the resulting contingency of even the most mature scientific results. Soler's picture of the development of science, which she believes the Gargamelle case supports but does not prove, is that at many,

many points in their research, scientists manage to establish some degree of stability, a kind of equilibrium, among the multitude of constraints and contingencies they face; yet we can easily imagine that their decisions at various points might well, and perfectly legitimately, have been different. Since these historical decisions become planks in the platform of ongoing science, their influence ramifies through all future work that depends upon them (here some generative entrenchment is at stake). The contingent nature of previous decisions and the reservations then felt are largely forgotten, concealed by later work, erased. Just as Stephen Jay Gould imagined that the tree of life on earth would look different each time the tape of biological evolution were replayed (a image that Pickering himself invokes in his chapter), so Soler imagines that the tree of scientific developments would look different than it does now, given our same universe and equally good science, if the tape of the historical evolution of science were replayed.

Soler takes the Gargamelle case to be emblematic of sophisticated scientific research. Insofar as this is true, her analysis appears to undercut the inference from robustness, and solidity more generally, to historical inevitability. A related issue is scientific realism. Beginning with Campbell, Levins and Wimsatt themselves, robustness and its relatives such as triangulation have often been made the basis of an argument for scientific realism, along the lines of the so-called miracle argument. According to this argument, were realism false, it would be a miracle that these independent lines of research should lead to exactly the same result. Soler argues here (and above in this introduction, Section 1.8) that the kind of flexibility-exploiting construction involved in making Wimsattian robustness claims possible in the first place undercuts this argument for strong realism. "(. . .) if there is a 'miracle' here, it seems to be of a different kind than the one involved in the realist argument". She refers to the perhaps surprising fact that scientists are so often able to establish the aforementioned stabilities within the Jamesian "blooming, buzzing confusion" of historical contingencies and constraints. But once we appreciate the amount of co-adjustment involved in somewhat forcing results to cohere, the sense of miracle begins to dissipate. As Kuhn put his version of the point, to some degree normal scientists have to "beat nature into line." This point feeds back into the contingency thesis.

B. **Frédéric Wieber's case study ("Multiple means of determination and multiple constraints on construction: robustness and strategies for modeling macromolecular objects")** examines the emergence and stabilization of a new scientific practice in protein chemistry, namely, a procedure developed in the 1960s and 1970s for modeling the structure of proteins. The problem is a formidable one, for proteins – those building blocks of all the life forms that we know – are exceedingly complex macromolecular objects, so complex that their structure must be described at four different, interacting levels: the amino acid sequence, the regular subunit arrangement (as in the alpha helix), the three-dimensional folding that is so crucial to chemical function, and the higher-order organization of multiple proteins (dimers, trimers, etc.). In principle, quantum theory can explain the bonding, protein folding, and so on; but the application of quantum theory was (and largely remains today) so far beyond our computational capability as to be completely intractable.

Given that human scientists are limited beings with limited resources, how did they go about tackling the problem of modeling proteins back in the 1960s? Why did the protein chemists develop, and how did they stabilize, a particular modeling strategy rather than pursue other options? What did it take to convince the community that a particular modeling procedure was reliable, and how well do the scientific practices involved fit Wimsatt's robustness analysis versus an alternative account of solidity?

Like Wimsatt himself, Wieber begins from population biologist Richard Levins' choice of modeling strategy from the 1960s, but Wieber eventually concludes that the solidity of the protein modeling procedure has more to do with Wimsatt's concept of generative entrenchment than with the robustness scheme. What is involved is "a mutual and iterative adjustment" of "the three limited resources" available to proteins scientists, namely "theoretical, empirical and technological" (especially computational) resources. Levins had argued that it is humanly scientifically impossible to provide population models of complex biological systems that are equally faithful to the demands of generality, realism and precision. Any model must partially sacrifice one of these dimensions. Levins also argued that robust results may nevertheless be obtained by considering multiple models (by varying parameters) and looking for convergence of results. As he famously summed up this strategy, "Our truth is the intersection of independent lies."

Since the investigation of protein structure in the 1960s and 1970s presented a similar challenge to scientists, Wieber uses Levins' framework as a tool to get a clearer understanding of the protein case. He shows why and how, in this case, scientists deliberately sacrificed the representational accuracy of their model to practical, computational imperatives, while nevertheless maintaining the hope that sufficient accuracy of the relevant predictions derived from the model would emerge from their work. Scientists used a patently inaccurate (by their own theoretical standards) formula that they had to feed with a high number of empirical parameters in order to build the model of one particular protein. This is the theoretical side of the model. Moreover, since the parameters are different from one molecule to the other, and since only a few of these parameters for a restricted number of proteins were experimentally determined (sometimes with discrepancies from one study to the other), scientists had to "exploit creatively" the existing empirical resources in order to estimate the unknown values. They did this by means of extrapolations and analogies from one kind of molecule to the other, and then, for each particular molecule, by adjusting the different parameters to one another in order to produce the so-called "force field" that (unlike single parameters in isolation) has chemical meaning. Next, these structural hypotheses had to be tested against the data. This is the empirical side of the modeling procedure. Finally, in order actually to use a given protein model, the scientists had to minimize the potential energy of the molecule, by intensively calculating a large number of conformations. Had computers not been available, the task would have been out of reach – which shows the historical dependency of this modeling procedure on the development of computer technology. This is the technological, computational side of the model.

Even with the new computational resources, the scientists remained almost overwhelmed by the complexity of protein chemistry. (This provides another illustration of Wimsatt's point that philosophers of science must treat human investigators (and communities of same) as limited beings, far from perfectly rational, let alone omniscient. Accordingly, they must make compromises and develop special methods for dealing with them. That is one of the reasons why the neglected issues of robustness and of solidity in general are so scientifically important and so philosophically interesting).

Wieber's case illustrates and clarifies the central role often played by computers in the solidification of a modeling procedure. In his protein case, recourse to computers greatly improved computational efficiency, but not only that. More fundamentally, it also modified the very content of the modeling procedure. For storing data in large databanks and partially black-boxing programs in the form of computer packages made available an increasingly large number of common parameters to more and more specialists. As a result, a more and more systematic and uniform process for the choice and estimation of the different parameters emerged and crystallized, replacing the human skill-dependent and more heterogeneous local solutions. Correlatively, an enlarged community of scientists began to use the procedure to interpret their experimental data about proteins in terms of molecular structures. This contributed to the refinement and optimization of the model parameters themselves through a back-and-forth movement between the experimental results obtained for more and more molecules on the one side, and the improvements introduced in the data base and computer program packages on the other side. As Wieber writes, "The computerization of the procedure of modeling is then really fundamental: with more and more models [of new proteins] constructed and effective calculations executed, scientists have been able to increasingly test the results produced against empirical data in order to *iteratively optimize* the parameters chosen for modeling."

1.9.5 Chapters 12, 13 and 14: Robustness, Scope, and Realism

The three last chapters widen the scope with respect to the robustness scheme *à la Wimsatt*. All three also question strong realism.

A. **Mieke Boon's chapter, "Understanding Scientific Practices: The Role of Robustness-Notions,"** deals with engineering sciences and, more broadly, with scientific research in the context of practical and technological applications. Boon is especially concerned with "the production and acceptance of physical phenomena as ontological entities; the role of instruments and experiments in their production; and the rule-like knowledge that is produced simultaneously".

Her aim is to understand what robustness might mean and how robustness is achieved in such scientific practices.

One overarching thesis of the whole analysis is that it is illegitimate to jump from robustness attributions to truth or reality attributions. "robustness, in the sense of multi-determination, cannot function as a truth-maker". What scientists achieve when they are successful is reliability. Hence a philosophy of science that cares about actual scientific practices must replace truth by reliability. Nevertheless, Boon endorses a minimal metaphysical realist belief, required, according to her, "in order to explain why scientific results can travel to other scientific fields or technological applications". This minimal realism implies the existence of an external, independent world that "stably sets limits" "to what we can *do* with it and to the regularities, causal relations, phenomena and objects that can possibly be determined". This realism is "minimal because it avoids the idea of a cognizable independent order or structure in the real world".

Boon distinguishes several "robustness-notions" that correspond to different uses of the term "robustness". According to a first, important use, "robust" is applied to the independent world as a whole. Here "robust" means "real", "stable" or the like. It points to an existing something which is supposed to be what it is once and for all, to resist us and to impose constraints on what we can do and think about it. This "robustness-notion" is a metaphysical category.

A second crucial robustness-notion that Mieke Boon applies to scientific practices points to the presupposition (when this sort of robustness obtains) that the same initial conditions will be followed by the same final conditions. Here, to say that our scientific practices are "robust" means that they are governed by the principle "Same conditions – same effects". This assumption has the status of a regulative principle: scientists cannot prove it, cannot "find out whether this principle is an empirical or metaphysical truth"; but practitioners indeed assume it and need it as a "condition of possibility" and as "a guiding principle" of any scientific research in the empirical sciences.

This latter assumption (robustness in the sense of "Same conditions – same effects") is related to the former assumption that there is an independent, real, stable world (robustness in the sense of reality and stability) imposing constraints on our experimental activities. But these regularities and reproducible connexions are not given as such, for they must be explored and constructed as technological achievements. "Experimental interventions with technological devices will (...) produce knowledge of conditions that are causally relevant to the reproducible production of a phenomenon described by $A \rightarrow B$, which is presented in 'rule-like' knowledge in the form: , unless (K and/or X).

How is this achieved? At this stage, a third robustness-notion (familiar to the readers of Wimsatt) enters the scene: multiple-derivability. Under the regulative idea of the robustness of scientific practices in the sense of "Same conditions – same effects", practitioners look for such kinds of practical recipes (my terminology). By repeating experiments, they vary their conditions and even the kinds of experiments. In this process they sometimes succeed in achieving a practical recipe of the kind "Same conditions – same effects". When this happens, the ‘something’ corresponding to the effects is said to be reproducible, stable, invariant. From this analysis of scientific practices, Boon puts forward a third robustness-notion: repetition and multiple-determination, which is a methodological category.

Now, the scientific results that, through this robust method, are built and recognized to be reproducible, stable and invariant, are also taken to be robust achievements in a different (although related) sense. In Boon’s terminology, when the scientific results involved are measured or observed physical occurrences, “robust” means “reproducible”, and when the scientific results involved are interpreted as phenomena described by $A \rightarrow B$, robust means “stable” and “invariant”. This is the fourth robustness-notion of Boon’s taxonomy. It is an ontological category, because when a result is reproducible, stable or invariant under multiple determinations, not only is it taken as acceptable, but also it acquires an ontological status.

In scientific practices, reproducibility, stability and invariance work as criteria for the acceptance that a ‘real something’ has been found. But that is still not all. As soon as an effect B has been recognized reproducible by the robust method of multi-determination, the rule-like knowledge of the form: “$A+C_{\text{device}}$, will produce the same effects, B, unless (K and/or X)” is also recognized to be reliable. Here we encounter the fifth and last of Boon’s robustness-notion. Here “robust” means “reliable”, applies to rule-like knowledge, and is an epistemological category. (This robustness notion also applies, in other ‘more theoretical’ contexts, to phenomenological laws, scientific models and even to fundamental theories – but in that latter case Boon prefers to talk about “empirical adequacy,” here borrowing van Fraassen’s expression).

These robustness-notions must be analytically distinguished, but in practice, the robustness attributions of the different kinds are essentially related and entangled: “Regulative, methodological, and epistemological or ontological criteria are used in a mutual interplay”. Robustness in the sense of reproducibility, stability and invariance works as an ontological criterion (a criterion for the acceptance of a phenomenon described by $A \rightarrow B$ as a real phenomenon: ontological robustness). It also works as an epistemological criterion (a criterion for the acceptance of the rule-like knowledge required for the (re)production of the phenomenon *as reliable* rules: epistemological robustness). These two kinds of robustness are essentially related to robustness in the sense of multi-determination (methodological robustness), since multiple-determination works as a methodological criterion for justifying the attributions of invariance (from which the ontological and reliable statuses are in turn attributed). Methodological robustness rests in turn on the regulative principle ‘Same conditions – same effects’, since the presupposition that our scientific practices are governed by such immutable regularities justifies the method of repetitions and variations as the way to delimitate such regularities. Finally, this regulative robustness is related to the metaphysical robustness, through the presupposition of the existence one real independent immutable world (a robust world in this sense) which imposes fixed constraints, and hence fixed regularities, to our experience and to what we can do and think about the world.

B. **Andrew Pickering's chapter, "The Robustness of Science and the Dance of Agency"**, is primarily driven by an ontologically-oriented interest rather than an epistemologically-oriented one. He understands the use of robustness terminology in Science Studies as a means to point to "the otherness of the world", with the intention to re-affirm and try to vindicate the existence of a non-human contribution of this world to our science. This use has to be situated in a given intellectual context. It expresses the attempt to find a viable middle way between two antagonist positions: on the one hand, the strongly counterintuitive relativist thesis that science is merely a social construction (understood as the claim that human beings can say whatever they want about the world, so that "the otherness of science vanishes"); and on the other hand, the equally strong and untenable realist thesis that our physical theories 'correspond' to a unique external world that is what it is once and for all, a conception in which scientists have no choice and science is "absolutely *other* to its producers and users".

So the question is: Does the world constitutively enter into science, how and in what sense? Or in other words: Is our science ontologically robust? Pickering's aim is to articulate his own position with respect to this "ontological sense of the robustness of science". To do so, Pickering first focuses on concrete actions and material aspects of science, convinced that "If there is a certain nonhuman toughness about scientific knowledge, it is grounded in performative (not cognitive) relations with the material world". But after having discussed the case of the first bubble chamber, he enlarges the scope: he examines the case of our conceptual knowledge about the material performances of instruments (in other words the case of experimental facts) and makes comments on the conceptual knowledge issuing from "purely conceptual practice".

At the end of the day, Pickering's answer to the initial question is the same for all of these kinds of scientific achievements, and it is positive: yes, ontologically robust are both our technological material achievements (the "free-standing machines and instruments" that "stand apart" from humans, "operate reliably" and reproduce the same performances independently of the individuals involved); and yes, ontologically robust is also our conceptual knowledge (experimental statements about physical phenomena or "purely conceptual systems"). In other words, the otherness of the world constitutively enters at all levels of our science.

Typically, the otherness of the world manifests itself, in some phases of research categorized as "phases of passivity", through material, instrumental performances. These manifestations are unpredictable and uncontrolled by humans. Practitioners do not know in advance what they will be. They have to wait for their otherness and to deal with it, and dealing with it, sometimes they succeed to obtain free standing-machines able to produce and reproduce stable material performances. This is "the primary sense in which the world enters constitutively into science – and the primary sense in which science is a robust enterprise and not a mere construction". Something similar holds for human attempts of conceptualizations: scientists cannot know in advance where this or that theoretical assumption will lead them. This motivates Pickering to nuance the metaphor of the plasticity of science he used in previous works: "at the conceptual as well as the material level the plasticity metaphor fails, precisely in that (. . .) scientists (. . .) have genuinely to find out what the upshot of that will be."

But if the otherness of the world constitutively enters at all levels of our science, the ontological robustness of our experimental recipes and scientific knowledge does not allow us to extract from this knowledge, and to contemplate apart of it, anything like a 'pure' bit of this external and independent world. The contribution of the world cannot be extracted and separated from the contribution of the human beings who make science and of the societies in which they evolve. The human and nonhuman contributions are irreducibly "mangled". "The material performance of instruments is indeed constitutive of the knowledge they produce, though prior scientific conceptualisations of the world are constitutive too, and this in an irrevocably intertwined fashion." Scientists deal with what Pickering elsewhere calls "symbioses", that is, complex structures made of irreducible intertwined human and nonhuman elements. All strive for "interactive stabilisations"; nobody knows in advance whether the material performances and the conceptual hypotheses will "fit together and interactively stabilise one another" and with the other scientific pieces already in place; and sometimes, good interactive stabilisations associated with a powerful "*machinic grip* on the world" are achieved. Here lies the manifestation of the otherness of the world and the robustness of science. But this otherness is not an absolute, "unsituated otherness", and this robustness does not mean any inevitability of the content of human knowledge.

However successful our scientific knowledge is, however satisfying is the symbiosis, however impressive is our machinic grip on the world and control on phenomena, we cannot "factor out the human side of the dance of agency": *it remains* "*our knowledge" and can never be equated with "something forced upon us by nature itself*" (emphasis added, LS). "We can indeed specify the source of science's robustness in dances of agency, especially with the material world, and in the production of free-standing machines and instruments", but the material performances always occur in the framework of some human questions, material and cultural resources, beliefs and values (all of which could have been different). The human reactions to these manifestations in the "active phases" which follow the passive ones are by no mean something that could be considered as pre-determined and unique. Different individuals or groups act differently in the same configuration and often favour different "accommodations", as the history of science so often shows. The material performances can be accommodated in different ways at each stage of the scientific development. And it happens, according to Pickering who gives examples, that several good inter-stabilisations and machinic grips are achieved, which all are *at the same time* robust *and* very different in content (sometimes even incommensurable).

So scientific development is an open, not pre-determined process, and scientific achievements of all kinds, although truly robust, are genuinely path-dependent and contingent. In the iterative process of passive and active phases, any kind of ingredient and dimension can be transformed and reconfigured: the individual scientists involved; the devices and their tangible performances; the ideas and practices of proof; the scientific forms of life; the idea of science and the place of science in our culture; as well as our ideas and commitments about the world and the kind of place the world is.

Pickering concludes that his account, by providing an alternative explanation, "somehow *defangs* realism and makes it a less pressing topic" and "undermine the intuition of uniqueness that goes along with" realism. The alternative 'symbiotic explanatory scheme' of solidity (as I would call it) is very general, if not universal (as Pickering writes: "the mangle is a sort of *theory of everything*"). It is assumed to apply to any kind of scientific practice (instrumental, experimental, purely theoretical, mixed kind of scientific practices). It is assumed to apply to all scales of scientific research (at the micro-, meso- and macro-levels). And it is moreover assumed to apply at an even more panoramic scale, since science, its knowledge and instruments have also to be in symbiosis with the rest of the society: with respect to different social values, our most beautiful theories could reduce to noise and our most efficient scientific instruments to "a pile of useless junk".

C. **Tom Nickles** opens his chapter, "**Dynamic Robustness and Design in Nature and Artifact**", by broadening the scope of robustness considerations to include apparently heterogeneous complex systems, both natural and artificial. Examples are large-scale industrial systems such as nuclear power plants and the electric power grid, information technological systems such as the Internet, but also epistemic systems such as a scientific specialty area and its products. Nickles is concerned with all "humanly constructed (explicitly or implicitly designed or engineered), evolved and evolving complex technological systems of *inquiry* and their products", especially "at innovative frontiers", where the systems in question are undergoing change at any of several levels (experimental, theoretical, instrumental, methodological, axiological). The stability of a Kuhnian paradigm could serve as an example, as we shall see.

Nickles' central question is whether we can normally hope to obtain an ever-increasing and cumulative robustness in either our epistemic systems or our complex material systems, up to something close to a zero-risk stage. And his answer is no: every attempt to increase robustness in one identified respect may, as far as we can usually know, create new fragilities in another respect. These new fragilities are often unexpected, unpredicted, in practice largely unpredictable, especially for innovative and bound-to-evolve systems. Moreover, the resulting failures can be worse than those prevented by the new robustness measures.

According to Nickles, this situation is inescapable because it is due to the very nature of the systems involved, especially owing to their complexity, high degree of connectivity, "extreme non-linearity" and designed-to-be-dynamic character.

Hence the new fragilities, accidents or failures subsequent to the robustness "improvements" should be considered as normal and endogenous, and not, as is commonly the case, as exogenous avoidable attacks from the outside or consequences of merely human errors.

Thus the "cumulative fragility-reduction thesis" or "convergent risk-reduction thesis is false when applied to epistemic systems". The hope that robustness analysis will gradually eliminate more and more errors and will bring us closer and closer to the ideal of invulnerability to major failures must be abandoned, even as a regulative ideal. Rather, we have to recognize and to prepare for "a direct coupling of robustness to fragility" and the inevitability of a "robustness-fragility tradeoff". "nothing that we can humanly do can prevent occasional, surprising avalanches of failure".

For the analysis of the robustness of epistemic systems and the argumentation of his "robustness-fragility tradeoff" thesis, Nickles exploits recent developments related to complex systems – all the more interesting for the completeness of a volume on robustness because these important resources are not represented elsewhere in the book.

Inspired by social scientist Charles Perrow's *Normal Accidents: Living with High-Risk Technologies* (1984), Nickles extends Perrow's thesis to epistemic systems while keeping also in view material and organizational networks such as the Internet and the commercial airline hub system. Perrow's qualitative risk analysis is complemented by more quantitative and technical literature on complex systems, e.g., network theory, the highly optimized tolerance (HOT) model of physicists Jean Carlson and John Doyle, and the study of system behaviors that can be described by distributions with so-called "heavy tails".

Heavy-tailed distributions are potentially relevant to robustness analysis because some such distributions signal the existence of a so-called "power law", and power law distributions have supposedly been found to characterize relevant aspects of systems of various kinds, including the distribution of failures. Insofar as this is true, it is frightening, because the incidence of failures does not drop off exponentially with the size of the failure as with a Gaussian distribution. (Earthquakes have a power law distribution.) Some investigators hope to discover mechanisms underlying these distributions that can be parlayed into a general trans-disciplinary "science of order and connectivity" or general complexity theory that would, among other things, help us understand robustness and fragility in a more theoretically sophisticated manner. Other experts are not so sanguine.

I should like to mention three other points from Nickles' chapter.

One that is very important to my eyes, but not often stressed, is that robustness must be relativized to specific kinds of potential threats or failures. It makes no sense to ask: Is a system robust (*tout court*)? We must specify: robust *with respect to what feature*? What do we want to avoid? For instance (and this also illustrates the tradeoff between robustness and fragility), an epistemic system of the axiomatic kind is robust to failures of logical entailment but is highly fragile to problems at the level of its starting principles, since any failure at this level will propagate instantaneously through the whole system and produce "a disastrous cascade of failure" and a collapse. Another example is this. (I choose this example since, at the end of the chapter, Nickles applies his ideas to Kuhn's model of science; although my illustration is freely inspired by his developments.) If the complex system is a scientific paradigm *à la Kuhn*, then, among the kinds of potential failure, we typically think, as Nickles stresses, to "empirical failure", that is, refuted predictions. But I think we could also count as a failure any apparently insurmountable barrier to fulfilling the main desiderata that the adherents of a given paradigm feel it crucial to satisfy, from Kuhn's "big five" to more specific requirements such as 'explain everything in mechanical terms only'. In sum, robustness is relative to various dimensions, in addition to be a matter of degree along each dimension.

Nickles also insists that "there is a prospective dimension of robustness" and that we should take into account "prospective robustness" in addition to the "purely retrospective conception of robustness". The retrospective conception assesses the robustness at a given point on the basis of what has happened up to now, examining the way the epistemic system has in fact resisted failures and has been fruitful. (In the case of scientific theories, this is often measured by philosophers in terms of the degree of justification at a given time.) But because "human designers of epistemic systems possess a degree of lookahead", retrospective assessments do not exhaust what human designers take into account when they evaluate the robustness of their epistemic systems. Their evaluations also incorporate judgments about "prospective fertility", intuitions and bets about the future fruitfulness of the system, its ability to resist to multiple threats and to fulfill the most important desiderata imposed on it. A robust system at the research frontier is one that possesses "a strong heuristic promise". This prospective dimension of robustness is "crucial for decision-making", especially when choices have to be made between several competing options, including research proposals. A robust system is a system that has proved to be robust up to now *and* that is perceived as promising future robustness. So both past-informed and future-oriented perspectives must be taken into account in evaluations of robustness.

Lastly, Nickles provides an interesting insight into ways in which his extension of Perrow's thesis and general theories of networks could help us to understand better some aspects of Kuhn's famous conception of mature science. In particular, this approach illuminates Kuhn's claims that the very nature of normal science "prepares the way for its own change" (in Kuhn's words). So here, too, we find the tradeoff put forward by Nickles between robustness and fragility, in this case in relation to scientific paradigms.

Acknowledgements Concerning the content of this introduction, I am grateful to Jacob Stegenga and Thomas Nickles for their useful comments. Many thanks also to them, and to Emiliano Trizio, for their corrections and suggestions of improvement concerning the English language.

More generally, my personal research on robustness has benefited from a collective project, called 'PratiScienS', which I initiated in 2007 in Nancy, France, and have led since that time. The aim of the PratiScienS group is to evaluate what we have learned about science from the practice turn in the studies devoted to science. The issue of robustness is one of the central axes of the project. I am grateful to the members of the group for fruitful exchanges on the subject.

The PratiScienS project is supported by the ANR (Agence Nationale de la Recherche), the MSH Lorraine (Maison des Sciences de l'Homme), the Région Lorraine, the LHSP – Laboratoire d'Histoire des Sciences et de Philosophie – Archives Henri Poincaré (UMR 7117 of the CNRS) and the University of Nancy 2. The support of these institutions enabled the PratiScienS group to organize, in June 2008 in Nancy, a conference on robustness to which many contributors of the present book participated.

References

Hacking, Ian. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, edited by A. Pickering, 29–64. Chicago and London: The University of Chicago Press.

Hacking, Ian. 1999. *The Social Construction of What*? Cambridge, MA: Harvard University Press.

Hacking, Ian. 2000. "How Inevitable are the Results of Successful Science?" *Philosophy of Science* 67:58–71.

Kuhn, Thomas. 1983. "Commensurability, Comparability, Communicability." In *Proceedings of the 1982 Biennal Meeting of the Philosophy of Science Association*, edited by P.D. Asquith and T. Nickles, 669–88. East Lansing, MI: Philosophy of Science Association.

Nederbragt, Hubertus. 2003. "Strategies to Improve the Reliability of a Theory: The Experiment of Bacterial Invasion into Cultured Epithelial Cells." *Studies in History and Philosophy of Biological and Biomedical Sciences* 34:593–614.

Perrow, Charles. 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Pickering, Andrew. 1984. *Constructing Quarks, a Sociological History of Particle Physics*. Chicago and London: The University of Chicago Press.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency and Science*. Chicago and London: The University of Chicago Press.

Pickering, Andrew. 201X. "Science, Contingency and Ontology." In *Science as It Could Have Been. Discussing the Contingent/Inevitable Aspects of Scientific Practices*, edited by L. Soler, E. Trizio, and A. Pickering. In progress.

Schindler, Samuel. 201X. *Weak Neutral Currents Revisited* (under review).

Soler, Léna. 2006a. "Contingence ou inévitabilité des résultats de notre science?" *Philosophiques* 33(2):363–78.

Soler, Léna. 2006b. "Une nouvelle forme d'incommensurabilité en philosophie des sciences?" *Revue philosophique de Louvain* 104(3):554–80.

Soler, Léna. 2008a. "The Incommensurability of Experimental Practices: The Incommensurability *of What*? An Incommensurability *of the Third-Type*?" In *Rethinking Scientific Change and Theory Comparison. Stabilities, Ruptures, Incommensurabilities*? edited by L. Soler, H. Sankey, and P. Hoyningen, 299–340. Dordrecht: Springer, Boston Studies for Philosophy of Science.

Soler, Léna. 2008b. "Are the Results of our Science Contingent or Inevitable? Introduction of a Symposium Devoted to the Contingency Issue." *Studies in History and Philosophy of Science* 39:221–29. Dordrecht: Springer, Boston Studies for Philosophy of Science.

Soler, Léna. 2008c. "Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue." *Studies in History and Philosophy of Science* 39:230–41. Dordrecht: Springer, Boston Studies for Philosophy of Science.

Soler, Léna. 201X. "A General Structural Argument in Favor of the Contingency of Scientific Results." In *Science as It Could Have Been. Discussing the Contingent/Inevitable Aspects of Scientific Practices*, edited by L. Soler, E. Trizio, and A. Pickering. In progress.

Van Bendegem, Jean-Paul, Amirouche Moktefi, Valéria Giardino, and Sandra Mols, eds. 2012. *From Practice to Results in Logic and Mathematics. Philosophia Scientiae,* Special Issue, 16(2), February 2012.

Wimsatt, William. 1981. “Robustness, Reliability and Overdetermination.” In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins, 125–63. San Francisco, CA: Jossey-Bass Publishers. Reprinted in (Wimsatt 2007a), 43–71.

Wimsatt, William. 2007a. *Re-engineering Philosophy for Limited Beings, Piecewise Approximations to Reality*. Cambridge, MA, and London, England: Harvard University Press.

Wimsatt, William. 2007b. Robustness and Entrenchment, How the Contingent Becomes Necessary. *Re-engineering Philosophy for Limited Beings, Piecewise Approximations to Reality*, Chapter 7, 133–45. Cambridge, MA and London, England: Harvard University Press.

Chapter 2
Robustness, Reliability, and Overdetermination (1981)

William C. Wimsatt

Philosophy ought to imitate the successful sciences in its methods, so far as to proceed only from tangible premises which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be so slender, provided they are sufficiently numerous and intimately connected
Peirce [1868] 1936, p. 141

Our truth is the intersection of independent lies
Levins 1966, p. 423

The use of multiple means of determination to "triangulate" on the existence and character of a common phenomenon, object, or result has had a long tradition in science but has seldom been a matter of primary focus. As with many traditions, it is traceable to Aristotle, who valued having multiple explanations of a phenomenon, and it may also be involved in his distinction between special objects of sense and common sensibles. It is implicit though not emphasized in the distinction between primary and secondary qualities from Galileo onward. It is arguably one of several conceptions involved in Whewell's method of the "consilience of inductions" (Laudan 1971) and is to be found in several places in Peirce.

Indeed, it is to be found widely among the writings of various scientists and philosophers but, remarkably, seems almost invariably to be relegated to footnotes, parenthetical remarks, or suggestive paragraphs that appear without warning and

From M. Brewer and B. Collins, eds., (1981); *Scientific Inquiry in the Social Sciences* (a festschrift for Donald T. Campbell), San Francisco: Jossey-Bass, pp. 123–162.

W.C. Wimsatt (✉)
Department of Philosophy and Conceptual and Historical Studies of Science, The University of Chicago, Chicago, IL, USA

Center for Philosophy of Science, University of Minnesota, St. Paul, MN, USA
e-mail: wwim@uchicago.edu

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_2,

vanish without further issue. While I will point to a number of different applications of multiple determination which have surfaced in the literature, Donald Campbell has done far more than anyone else to make multiple determination a central focus of his work and to draw a variety of methodological, ontological, and epistemological conclusions from its use (see Campbell 1958, 1966, 1969a, 1977; Campbell and Fiske 1959; Cook and Campbell 1979). This theme is as important a contribution as his work on evolutionary epistemology; indeed, it must be a major unappreciated component of the latter: multiple determination, because of its implications for increasing reliability, is a fundamental and universal feature of sophisticated organic design and functional organization and can be expected wherever selection processes are to be found.

Multiple determination—or *robustness,* as I will call it—is not limited in its relevance to evolutionary contexts, however. Because of its multiplicity of uses, it is implicit in a variety of criteria, problem-solving procedures, and cognitive heuristics which have been widely used by scientists in different fields, and is rich in still insufficiently studied methodological and philosophical implications. Some of these I will discuss, some I will only mention, but each contains fruitful directions for future research.

2.1 Common Features of Concepts of Robustness

The family of criteria and procedures which I seek to describe in their various uses might be called *robustness analysis.* They all involve the following procedures:

1. To analyze a *variety* of *independent* derivation, identification, or measurement processes.
2. To look for and analyze things which are *invariant* over or *identical* in the conclusions or results of these processes.
3. To determine the *scope* of the processes across which they are invariant and the *conditions* on which their invariance depends.
4. To analyze and explain any relevant *failures of invariance.*

I will call things which are invariant under this analysis "robust," extending the usage of Levins (1966, p. 423), who first introduced me to the term and idea and who, after Campbell, has probably contributed most to its analysis (see Levins 1966, 1968).

These features are expressed in very general terms, as they must be to cover the wide variety of different practices and procedures to which they apply. Thus, the different processes in clause 1 and the invariances in clause 2 may refer in different cases to any of the following:

a. Using different sensory modalities to detect the same property or entity (in the latter case by the detection of spatiotemporal boundaries which are relatively invariant across different sensory modalities) (Campbell 1958, 1966).
b. Using different experimental procedures to verify the same empirical relationships or generate the same phenomenon (Campbell and Fiske 1959).
c. Using different assumptions, models, or axiomatizations to derive the same result or theorem (Feynman 1965; Levins 1966; Glymour 1980).
d. Using the agreement of different tests, scales, or indices for different traits, as measured by different methods, in ordering a set of entities as a criterion for the "validity" (or reality) of the constructed property (or "construct") in terms of which the orderings of entities agree (Cronbach and Meehl 1955; Campbell and Fiske 1959).
e. Discovering invariance of a macrostate description, variable, law, or regularity over different sets of microstate conditions, and also determining the microstate conditions under which these invariances may fail to hold (Levins 1966, 1968; Wimsatt 1976a, b, 1980b).
f. Using matches and mismatches between theoretical descriptions of the same phenomenon or system at different levels of organization, together with Leibniz's law (basically that if two things are identical, no mismatches are allowed), to generate new hypotheses and to modify and refine the theories at one or more of the levels (Wimsatt 1976a, b, 1979).
g. Using failures of invariance or matching in a through f above to calibrate or recalibrate our measuring apparatus (for a, b or f) or tests (for d), or to establish conditions (and limitations on them) under which the invariance holds or may be expected to fail, and (for all of the above) to use this information to guide the search for explanations as to why the invariances should hold or fail (Campbell 1966, 1969a; Wimsatt 1976a, b).
h. Using matches or mismatches in different determinations of the value of theoretical parameters to test and confirm or infirm component hypotheses of a complex theory (Glymour 1980) and, in a formally analogous manner, to test and localize faults in integrated circuits.

One may ask whether any set of such diverse activities as would fit all these items and as exemplified in the expanded discussion below are usefully combined under the umbrella term *robustness analysis*. I believe that the answer must be yes, for two reasons. First, all the variants and uses of robustness have a common theme in the distinguishing of the real from the illusory; the reliable from the unreliable; the objective from the subjective; the object of focus from artifacts of perspective; and, in general, that which is regarded as ontologically and epistemologically trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting. The variations of use of these procedures in different applications introduce different variant tools or consequences which issue from this core theme and are explicable in terms of it. Second, all these procedures require at least partial *independence* of the various processes across which invariance is shown. And each of them is subject to a kind of systematic error leading to a kind of *illusory robustness* when we are led, on less than definitive evidence, to presume independence and our presumption turns out to be incorrect. Thus, a broad class of fallacious inferences in science can be understood and analyzed as a kind of failure of robustness.

Nonetheless, the richness and variety of these procedures require that we go beyond this general categorization to understand robustness. To understand fully the variety of its applications and its central importance to scientific methodology, detailed case studies of robustness analysis are required in each of the areas of science and philosophy where it is used.

2.2 Robustness and the Structure of Theories

In the second of his popular lectures on the character of physical law, Feynman (1965) distinguishes two approaches to the structure of physical theory: the Greek and the Babylonian approaches. The Greek (or Euclidean) approach is the familiar axiomatic one in which the fundamental principles of a science are taken as axioms, from which the rest are derived as theorems. There is an established order of importance, of ontological or epistemological priority, from the axioms out to the farthest theorems. The "Greek" theorist achieves postulational economy or simplicity by making only a small number of assumptions and deriving the rest—often reducing the assumptions, in the name of simplicity or elegance, to the minimal set necessary to derive the given theorems. The "Babylonian," in contrast, works with an approach that is much less well ordered and sees a theoretical structure that is much more richly connected:

> So the first thing we have to accept is that even in mathematics you can start in different places. If all these various theorems are interconnected by reasoning there is no real way to say "These are the most fundamental axioms," because if you were told something different instead you could also run the reasoning the other way. It is like a bridge with lots of members, and it is overconnected; if pieces have dropped out you can reconnect it another way. The mathematical tradition of today is to start with some particular ideas which are chosen by some kind of convention to be axioms, and then to build up the structure from there. What I have called the Babylonian idea is to say, "I happen to know this, and I happen to know that, and maybe I know that; and I work everything out from there. Tomorrow I may forget that this is true, but remember that something else is true, so I can reconstruct it all again. I am never quite sure of where I am supposed to begin or where I am supposed to end. I just remember enough all the time so that as the memory fades and some of the pieces fall out I can put the thing back together again every day" (Feynman 1965, pp. 46–47).

This rich connectivity has several consequences for the theoretical structure and its components. First, as Feynman (1965, pp. 54–55) observes, most of the fundamental laws turn out to be characterizable and derivable in a variety of different ways from a variety of different assumptions: "One of the amazing characteristics of nature is the variety of interpretational schemes which is possible. It turns out that it is only possible because the laws are just so, special and delicate. . . . If you modify the laws much you find that you can only write them in fewer ways. I always find that mysterious, and I do not understand the reason why it is that the correct laws of physics seem to be expressible in such a tremendous variety of ways. They seem to be able to get through several wickets at the same time." Although Feynman nowhere explicitly says so, his own choice of examples and other considerations that will emerge later suggest another ordering principle for fundamentality among laws of nature: *The more fundamental laws will be those that are independently derivable in a larger number of ways*. I will return to this suggestion later.

Second, Feynman also observes that this multiple derivability of physical laws has its advantages, for it makes the overall structure much less prone to collapse:

> At present we believe that the laws of physics have to have the local character and also the minimum principle, but we do not really know. If you have a structure that is only partly accurate, and something is going to fail, then if you write it with just the right axioms maybe only one axiom fails and the rest remain, you need only change one little thing. But if you write it with another set of axioms they may all collapse, because they all lean on that one thing that fails. We cannot tell ahead of time, without some intuition, which is the best way to write it so that we can find out the new situation. We must always keep all the alternative ways of looking at a thing in our heads; so physicists do Babylonian mathematics, and pay but little attention to the precise reasoning from fixed axioms (Feynman 1965, p. 54).

This multiple derivability not only makes the overall structure more reliable but also has an effect on its individual components. Those components of the structure which are most insulated from change (and thus the most probable foci for continuity through scientific revolutions) are those laws which are most robust and, on the above criterion, most fundamental. This criterion of fundamentality would thus make it natural (though by no means inevitable) that the most fundamental laws would be the least likely to change. *Given that different degrees of robustness ought to confer different degrees of stability, robustness ought to be a promising tool for analyzing scientific change.* Alternatively, the analysis of different degrees of change in different parts of a scientific theory may afford a way of detecting or measuring robustness.

I wish to elaborate and illustrate the force of Feynman's remarks arguing for the Babylonian rather than the Greek or Euclidean approach by some simple considerations suggested by the statistical theory of reliability. (For an excellent review of work in reliability theory, see Barlow and Proschan 1975, though no one has, to my knowledge, applied it in this context.)

A major rationale for the traditional axiomatic view of science is to see it as an attempt to make the structure of scientific theory as reliable as possible by starting with, as axioms, the minimal number of assumptions which are as certain as possible and operating on them with rules which are as certain as possible (deductive rules which are truth preserving). In the attempt to secure high reliability, the focus is on total elimination of error, not on recognizing that it will occur and on controlling its effects: it is a structure in which, if no errors are introduced in the assumptions and if no errors are made in choosing or in applying the rules, no errors will occur. No effort is spared in the attempt to prevent these kinds of errors from occurring. But it does not follow that this is the best structure for dealing with errors (for example, by minimizing their effects or making them easier to find) if they do occur. In fact, it is not. To see how well it handles errors that do occur, let us try to model the effect of applying the Greek or Euclidean strategy to a real (error-prone) theory constructed and manipulated by real (fallible) operators.

For simplicity, assume that any operation, be it choosing an assumption or applying a rule, has a small but finite probability of error, p_0. (In this discussion, I will assume that the probability of error is constant across different components and operations. Qualitatively similar results obtain when it is not.) Consider now the deductive derivation of a theorem requiring m operations. If the probabilities of failure in these operations are independent, then the probability of a successful derivation is just the product of the probabilities of success, $(1-p_0)$, at each operation. Where p_s stands for the probability of failing at this complex task (p_s because this is a serial task), then we have for the probability of success, $(1-p_s)$:

$$(1-p_s) = (1-p_0)^m$$

No matter how small p_0 is, as long as it is finite, longer serial deductions (with larger values of m) have monotonically decreasing probabilities of successful completion, approaching zero in the limit. *Fallible thinkers should avoid long serial chains of reasoning*. Indeed, we see here that the common metaphor for deductive reasoning as a chain is a poor one for evaluating probability of failure in reasoning. Chains always fail at their weakest links, chains of reasoning only most probably so.

When a chain fails, the release in tension protects other parts of the chain. As a result, failures in such a chain are not independent, since the occurrence of one failure prevents other failures. In this model, however, we are assuming that failures are independent of each other, and we are talking about probability of failure rather than actual failure. These differences result in a serious disanalogy with the metaphor of the argument as a chain. A chain is only as strong as the weakest link, but it is that strong; and one often hears this metaphor as a rule given for evaluating the reliability of arguments (see, for example, the quote from C.S. Peirce that begins this chapter). But a chain in which failure could occur at any point is always weaker than (in that it has a higher probability of failure than) its weakest link, except if the probability of failure everywhere else goes to zero. This happens when the weakest link in a chain breaks, but not when one link in an argument fails.

Is there any corrective medicine for this cumulative effect on the probability of error, in which small probabilities of error in even very reliable components cumulatively add up to almost inevitable failure? Happily there is. *With independent alternative ways of deriving a result, the result is always surer than its weakest derivation.* (Indeed, it is always surer than its *strongest* derivation.) This mode of organization—with independent alternative modes of operation and success if any one works—is parallel organization, with its probability of failure, p_p. Since failure can occur if and only if each of the m independent alternatives fails (assume, again, with identical probabilities p_0):

$$p_p = p_0^m$$

But p_0 is presumably always less than 1; thus, for $m > 1$, p_p is always less than p_0. Adding alternatives (or redundancy, as it is often called) always increases reliability, as von Neumann (1956) argued in his classic paper on building reliable automata with unreliable components. Increasing reliability through parallel organization is a fundamental principle of organic design and reliability engineering generally. It works for theories as well as it does for polyploidy, primary metabolism, predator avoidance, microprocessor architecture, Apollo moon shots, test construction, and the structure of juries.

Suppose we start, then, with a Babylonian (or Byzantine?) structure—a multiply connected, poorly ordered scientific theory having no principles singled out as axioms, containing many different ways of getting to a given conclusion and, because of its high degree of redundancy, relatively short paths to it (see Feynman 1965, p. 47)—and let it be redesigned by a Euclidean. In the name of elegance, the Euclidean will look for a small number of relatively powerful assumptions from which the rest may be derived. In so doing, he will eliminate redundant assumptions. The net effects will be twofold: (1) With a smaller number of assumptions taken as axioms, the mean number of steps in a derivation will increase, and can do so exponentially. This increased length of seriation will decrease reliability along any path to the conclusion. (2) Alternative or parallel ways of getting to a given conclusion will be substantially decreased as redundant assumptions are removed, and this decrease in "parallation" will also decrease the total reliability of the conclusion.

Each of these changes increases the unreliability of the structure, and both of them operating together produce a cumulative effect—if errors are possible, as I have supposed. Not only is the probability of failure of the structure greater after it has been Euclideanized, but the consequences of failure become more severe: with less redundancy, the failure of any given component assumption is likely to infirm a larger part of the structure. I will elaborate on this point shortly. It has not been studied before now (but see Glymour 1980) because of the dominance of the Cartesian Euclidean perspective and because of a key artifact of first-order logic.

Formal models of theoretical structures characteristically start with the assumption that the structures contain no inconsistencies. As a normative ideal, this is fine; but as a description of real scientific theories, it is inadequate. Most or all scientific theories with which I am familiar contain paradoxes and inconsistencies, either between theoretical assumptions or between assumptions and data in some combination. (Usually these could be resolved if one knew which of several eminently plausible assumptions to give up, but each appears to have strong support; so the assumptions—and the inconsistencies—remain.) This feature of scientific theories has not until now (with the development of nonmonotonic logic) been modeled, because of the fear of total collapse. In first-order logic, anything whatsoever follows from a contradiction; so systems which contain contradictions are regarded as useless.

But the total collapse suggested by first-order logic (or by highly Euclidean structures with little redundancy) seems not to be a characteristic of scientific theories. The thing that is remarkable about scientific theories is that the inconsistencies are walled off and do not appear to affect the theory other than very locally—for things very close to and strongly dependent on one of the conflicting assumptions. Robustness provides a possible explanation, perhaps the best explanation, for this phenomenon.

When an inconsistency occurs, results which depend on one or more of the contradictory assumptions are infirmed. This infection is transitive; it passes to things that depend on these results, and to their logical descendants, like a string of dominoes—until we reach something that has independent support. The independent support of an assumption sustains it, and the collapse propagates no further. If all deductive or inferential paths leading from a contradiction pass through robust results, the collapse is bounded within them, and the inconsistencies are walled off from the rest of the network. For each robust result, one of its modes of support is destroyed; but it has others, and therefore the collapse goes no further. Whether this is the only mechanism by which this isolation of contradictions could be accomplished, I do not know, but it is a possible way, and scientific constructs do appear to have the requisite robustness. (I am not aware that anyone has tried to formalize or to simulate this, though Stuart A. Kauffman's work on "forcing structures" in binary, Boolean switching networks seems clearly relevant. See, for example, Kauffman 1971, where these models are developed and applied to gene control networks.)

2.3 Robustness, Testability, and the Nature of Theoretical Terms

Another area in which robustness is involved (and which is bound to see further development) is Clark Glymour's account of testing and evidential relations in theories. Glymour argues systematically that parts of a theoretical structure can be and are used to test other parts of the theory, and even themselves. (His name for this is bootstrapping.) This testing requires the determination of values for quantities of the theory in more than one way: "If the data are consistent with the theory, then these different computations must agree [within a tolerable experimental error] in the value they determine for the computed quantity; but if the data are inconsistent with the theory, then different computations of the same quantity may give different results. Further and more important, what quantities in a theory may be computed from a given set of initial data depends both on the initial data and on the structure of the theory" (Glymour 1980, p. 113).

Glymour argues later (pp. 139–140) that the different salience of evidence to different hypotheses of the theory requires the use of a variety of types of evidence to test the different component hypotheses of the theory. Commenting on the possibility that one could fail to locate the hypothesis whose incorrectness is producing an erroneous determination of a quantity or, worse, mislocating the cause of the error, he claims: "The only means available for guarding against such errors is to have a variety of evidence so that as many hypotheses as possible are tested in as many different ways as possible. What makes one way of testing relevantly different from another is that the hypotheses used in one computation are different from the hypotheses used in the other computation. Part of what makes one piece of evidence relevantly different from another piece of evidence is that some test is possible from the first that is not possible from the second, or that, in the two cases, there is some difference in the precision of computed values of theoretical quantities" (Glymour 1980, p. 140).

A given set of data and the structure of the theory permit a test of a hypothesis (or the conjunction of a group of hypotheses) if and only if they permit determination of all of the values in the tested entity in such a way that contradictory determinations of at least one of these values could result (in the sense that it is not analytically ruled out). This requires more than one way of getting at that value (see Glymour 1980, p. 307). To put it in the language of the present chapter, *only robust hypotheses are testable.* Furthermore, a theory in which most components are multiply connected is a theory whose faults are relatively precisely localizable. Not only do errors not propagate far, but we can find their source quickly and evaluate the damage and what is required for an adequate replacement. If this sounds like a design policy for an automobile, followed to facilitate easy diagnostic service and repair, I can say only that there is no reason why our scientific theories should be less well designed than our other artifacts.

The same issues arise in a different way in Campbell's discussions (Campbell and Fiske 1959; Campbell 1969a, b, 1977; Cook and Campbell 1979) of single or definitional versus multiple operationalism. Definitional operationalism is the view that philosophers know as operationalism, that the meaning of theoretical terms is to be defined in terms of the experimental operations used in measuring that theoretical quantity. Multiple means of determining such a quantity represents a paradox for this view—an impossibility, since the means is definitive of the quantity, and multiple means multiple quantities. Campbell's multiple operationalism is not operationalism at all in this sense but a more tolerant and eclectic empiricism, for he sees the multiple operations as contingently associated with the thing measured. Being contingently associated, they cannot have a definitional relation to it; consequently, there is no barrier to accepting that one (robust) quantity has a number of different operations to get at it, each too imperfect to have a definitional role but together triangulating to give a more accurate and complete picture than would be possible from any one of them alone.

Campbell's attack on definitional operationalism springs naturally from his fallibilism and his critical realism. Both of these forbid a simple definitional connection between theoretical constructs and measurement operations: "One of the great weaknesses in definitional operationalism as a description of best scientific practice was that it allowed no formal way of expressing the scientist's prepotent awareness of the imperfection of his measuring instruments and his prototypic activity of improving them" (Campbell 1969a, p. 15). For a realist the connection between any measurement and the thing measured involves an often long and indirect causal chain, each link of which is affected and tuned by other theoretical parameters. The aim is to make the result insensitive to or to control these causally relevant but semantically irrelevant intermediate links: "What the scientist does in practice is to design the instrument so as to minimize and compensate for the stronger of these irrelevant forces. Thus, the galvanometer needle is as light as possible, to minimize inertia. It is set on jeweled bearings to minimize friction. It may be used in a lead-shielded and degaussed room. Remote influences are neglected because they dissipate at the rate of l/d2, and the weak and strong nuclear forces dissipate even more rapidly. But these are practical minimizations, recognizable on theoretical grounds as incomplete" (1969a, pp. 14–15).

The very same indirectness and fallibility of measurement that rule out definitional links make it advantageous to use multiple links: "[W] e have only *other invalid measures* against which to validate our tests; we have no 'criterion' to check them against. . . . A theory of the interaction of two theoretical parameters must be tested by imperfect exemplifications of each. . . . In this predicament, great inferential strength is added when each theoretical parameter is exemplified in 2 or more ways, each mode being as independent as possible of the other, as far as the theoretically irrelevant components are concerned. This general program can be designated *multiple operationalism*" (Campbell 1969a, p. 15).

Against all this, then, suppose one did have only one means of access to a given quantity. Without another means of access, even if this means of access were not made definitional, statements about the value of that variable would not be independently testable. Effectively, they would be as if defined by that means of access. And since the variable was not connected to the theory in any other way, it would be an unobservable, a fifth wheel: anything it could do could be done more directly by its operational variable. It is, then, in Margenau's apt phrase, a peninsular concept (Margenau 1950, p. 87), a bridge that leads to nowhere.

Philosophers often misleadingly lump this "peninsularity" and the existence of extra axioms permitting multiple derivations together as redundancy. The implication is that one should be equally disapproving of both. Presumably, the focus on error-free systems leads philosophers to regard partially identical paths (the paths from a peninsular concept and from its "operational variable" to any consequence accessible from either) and alternative independent paths (robustness, bootstrapping, or triangulation) as equivalent—because they are seen as equally dispensable if one is dealing with a system in which errors are impossible. But if errors are possible, the latter kind of redundancy can increase the reliability of the conclusion; the former cannot.

A similar interest in concepts with multiple connections and a disdain for the trivially analytic, singly or poorly connected concept is to be found in Putnam's (1962) classic paper "The Analytic and the Synthetic." Because theoretical definitions are multiply connected law-cluster concepts, whose meaning is determined by this multiplicity of connections, Putnam rejects the view that such definitions are stipulative or analytic. Though for Putnam it is theoretical connections, rather than operational ones, which are important, he also emphasizes the importance of a multiplicity of them: "Law-cluster concepts are constituted not by a bundle of properties as are the typical general names [cluster concepts] like 'man' and 'crow,' but by a cluster of laws which, as it were, determine the identity of the concept. The concept 'energy' is an excellent sample. . . . It enters into a great many laws. It plays a great many roles, and these laws and inference roles constitute its meaning collectively, not individually. I want to suggest that most of the terms in highly developed sciences are law-cluster concepts, and that one should always be suspicious of the claim that a principle whose subject term is a law-cluster concept is analytic. The reason that it is difficult to have an analytic relationship between law-cluster concepts is that . . . any one law can be abandoned without destroying the identity of the law-cluster concept involved" (p. 379).

Statements that are analytic are so for Putnam because they are singly connected, not multiply connected, and thus trivial: "Thus, it cannot 'hurt' if we decide always to preserve the law 'All bachelors are unmarried' . . . because bachelors are a kind of synthetic class. They are a 'natural kind' in Mill's sense. They are rather grouped together by ignoring all aspects except a single legal one. One is simply not going to find any . . . [other] laws about such a class" (p. 384).

Thus, the robustness of a concept or law—its multiple connectedness within a theoretical structure and (through experimental procedures) to observational results—has implications for a variety of issues connected with theory testing and change, with the reliability and stability of laws and the component parts of a theory, with the discovery and localization of error when they fail, the analytic-synthetic distinction, and accounts of the meaning of theoretical concepts. But these issues have focused on robustness in existing theoretical structures. It is also important in discovery and in the generation of new theoretical structures.

2.4 Robustness, Redundancy, and Discovery

For the complex systems encountered in evolutionary biology and the social sciences, it is often unclear what is fundamental or trustworthy. One is faced with a wealth of partially conflicting, partially complementary models, regularities, constructs, and data sets with no clear set of priorities for which to trust and where to start. In this case particularly, processes of validation often shade into processes of discovery—since both involve a winnowing of the generalizable and the reliable from the special and artifactual. Here too robustness can be of use, as Richard Levins suggests in the passage which introduced me to the term:

> Even the most flexible models have artificial assumptions. There is always room for doubt as to whether a result depends on the essentials of a model or on the details of the simplifying assumptions. This problem does not arise in the more familiar models, such as the geographical map, where we all know that contiguity on the map implies contiguity in reality, relative distances on the map correspond to relative distances in reality, but color is arbitrary and a microscopic view of the map would only show the fibers of the paper on which it is printed. But in the mathematical models of population biology, it is not always obvious when we are using too high a magnification.
>
> Therefore, we attempt to treat the same problem with several alternative models, each with different simplifications, but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence, our truth is the inter section of independent lies (Levins 1966, p. 423).

Levins is here making heuristic use of the philosopher's criterion of logical truth as true in all possible worlds. He views robustness analysis as "sampling from a space of possible models" (1968, p. 7). Since one cannot be sure that the sampled models are representative of the space, one gets no guarantee of logical truth but, rather, a heuristic (fallible but effective) tool for discovering empirical truths which are relatively free of the details of the various specific models.

Levins talks about the robustness of theorems or phenomena or consequences of the models rather than about the robustness of the models themselves. This is necessary, given his view that any single model makes a number of artifactual (and therefore nonrobust) assumptions. A theory would presumably be a conceptual structure in which many or most of the fundamental theorems or axioms are relatively robust, as is suggested by Levins' statement (1968, p. 7) "A theory is a cluster of models, together with their robust consequences."

If a result is robust over a range of parameter values in a given model or over a variety of models making different assumptions, this gives us some independence of knowledge of the exact structure and parameter values of the system under study: a prediction of this result will remain true under a variety of such conditions and parameter values. This is particularly important in scientific areas where it may be difficult to determine the parameter values and conditions exactly.

Robust theorems can thus provide a more trustworthy basis for generalization of the model or theory and also, through their independence of many exact details, *a sounder basis for predictions from it.* Theory generalization is an important component of scientific change, and thus of scientific discovery.

Just as robustness is a guide for discovering trustworthy results and generalizations of theory, and distinguishing them from artifacts of particular models, it helps us to distinguish signal from noise in perception generally. Campbell has furnished us with many examples of the role of robustness and pattern matching in visual perception and its analogues, sonar and radar. In an early paper, he described how the pattern and the redundancy in a randomly pulsed radar signal bounced off Venus gave a new and more accurate measurement of the distance to that planet (Campbell 1966).

The later visual satellite pictures of Mars and its satellite Deimos have provided an even more illuminating example, again described by Campbell (1977) in the unpublished William James Lectures (lecture 4, pp. 89 and 90). The now standard procedures of image enhancement involve combining a number of images, in which the noise, being random, averages out; but the signal, weak though usually present, adds in intensity until it stands out. The implicit principle is the same one represented explicitly in von Neumann's (1956) use of "majority organs" to filter out error: the combination of parallel or redundant signals with a threshold, in which it is assumed that the signal, being multiply represented, will usually exceed threshold and be counted; and the noise, being random, usually will fall below threshold and be lost. There is an art to designing the redundancy so as to pick up the signal and to setting the threshold so as to lose the noise. It helps, of course, if one knows what he is looking for. In this case of the television camera centered on Mars, Deimos was a moving target and—never being twice in the same place to add appropriately (as were the static features of Mars)—was consequently filtered out as noise. But since the scientists involved knew that Deimos was there, they were able to fix the image enhancement program to find it. By changing the threshold (so that Deimos and some noise enter as—probably smeared—signal), changing the sampling rate or the integration area (stopping Deimos at the effectively same place for two or more times), or introducing the right kind of spatiotemporal correlation function (to track Deimos's periodic moves around Mars), could restore Deimos to the pictures again. Different tunings of the noise filters and different redundancies in the signal were exploited to bring static Mars and moving Deimos into clear focus.

We can see exactly analogous phenomena in vision if we look at a moving fan or airplane propeller. We can look through it (filtering it out as noise) to see something behind it. Lowering our threshold, we can attend to the propeller disk as a colored transparent (smeared) object. Cross-specific variation in flickerfusion frequency indicates different sampling rates, which are keyed to the adaptive requirements of the organism (see Wimsatt 1980a, pp. 292–297). The various phenomena associated with periodic stroboscopic illumination (apparent freezing and slow rotation of a rapidly spinning object) involve detection of a lagged correlation. Here, too, different tunings pick out different aspects of or entities in the environment. This involves a use of different heuristics, a matter I will return to later.

I quoted Glymour earlier on the importance of getting the same answer for the value of quantities computed in two different ways. What if these computations or determinations do not agree? The result is not always disastrous; indeed, when such mismatches happen in a sufficiently structured situation, they can be very productive.

This situation could show that we were wrong in assuming that we were detecting or determining the same quantity; but (as Campbell 1966, was the first to point out), if we assume that we *are* determining the same quantity but "through a glass darkly," the mismatch can provide an almost magical opportunity for discovery. Given imperfect observations of a thing-we-know-not-what, using experimental apparatus with biases-we-may-not-understand, we can achieve both a better understanding of the object (it must be, after all, that one thing whose properties can produce these divergent results in these detectors) and of the experimental apparatus (which are, after all, these pieces that can be affected thus divergently by this one thing).

The constraint producing the information here is the identification of the object of the two or more detectors. If two putatively identical things are indeed identical, then any property of one must be a property of the other. We must resolve any apparent differences either by giving up the identification or locating the differences not in the thing itself but in the interactions of the thing with different measuring instruments. And this is where we learn about the measuring instruments. Having then acquired a better knowledge of the biases of the measuring instruments, we are in a better position not only to explain the differences but also, in the light of them, to give a newly refined estimate of the property of the thing itself. This procedure, a kind of "means-end" analysis (Wimsatt 1976a; Simon 1969) has enough structure to work in any given case only because of the enormous amount of background knowledge of the thing and the instruments which we bring to the situation. What we can learn (in terms of localizing the source of the differences) is in direct proportion to what we already know.

This general strategy for using identifications has an important subcase in reductive explanation. I have argued extensively (Wimsatt 1976a, part II, 1976b, 1979) that the main reason for the productiveness of reductive explanation is that interlevel identifications immediately provide a wealth of new hypotheses: each property of the entity as known at the lower level must be a property of it as known at the upper level, and conversely; and usually very few of these properties from the other level have been predicated of the common object. The implications of these predictions usually have fertile consequences at both levels, and even where the match is not exact, there is often enough structure in the situation to point to a revised identification, with the needed refinements. This description characterizes well the history of genetics, both in the period of the localization of the genes on chromosomes (1883–1920) and in the final identification of DNA as the genetic material (1927–1953). (For the earlier period see, for example, Allen 1979; Moore 1972; Darden 1974; Wimsatt 1976a, part II. For the later period see Olby 1974.) Indeed, the overall effect of these considerations is to suggest that *the use of identities in a structured situation for the detection of error may be the most powerful heuristic known and certainly one of the most effective in generating scientific hypotheses.*

Also significant in the connection between robustness and discovery is Campbell's (1977) suggestion that things with greater entitativity (things whose boundaries are more robust) ought to be learned earlier. He cites suggestive support from language development for this thesis, which Quine's (1960) views also tend to support. I suspect that robustness could prove to be an important tool in analyzing not only what is discovered but also the order in which things are discovered.

There is some evidence from work with children (Omanson 1980a, b) that components of narratives which are central to the narrative, in that they are integrated into its causal and its purposive or intentional structure, are most likely to be remembered and least likely to be abstracted out in summaries of the story. This observation is suggestively related both to Feynman's (1965, p. 47) remark quoted above, relating robustness to forgetting relationships in a multiply connected theory, and to Simon's (1969) concept of a blackboard work space, which is maintained between successive attempts to solve a problem and in which the structure of the problem representation and goal tree may be subtly changed through differential forgetting. These suggest other ways in which robustness could affect discovery processes through differential effects on learning and forgetting.

2.5 Robustness, Objectification, and Realism

Robustness is widely used as a criterion for the reality or trustworthiness of the thing which is said to be robust. The boundaries of an ordinary object, such as a table, as detected in different sensory modalities (visually, tactually, aurally, orally), roughly coincide, making them robust; and this is ultimately the primary reason why we regard perception of the object as veridical rather than illusory (see Campbell 1958, 1966). It is a rare illusion indeed which could systematically affect all of our senses in this consistent manner. (Drug induced hallucinations and dreams may involve multimodal experience but fail to be consistent through time for a given subject, or across observers, thus failing at a higher level to show appropriate robustness.)

Our concept of an object is of something which exemplifies a multiplicity of properties within its boundaries, many of which change as we move across its boundary. A one-dimensional object is a contradiction in terms and usually turns out to be a disguised definition—a legal or theoretical fiction. In appealing to the robustness of boundaries as a criterion for objecthood, we are appealing to this multiplicity of properties (different properties detected in different ways) and thus to a time-honored philosophical notion of objecthood.

Campbell (1958) has proposed the use of the coincidence of boundaries under different means of detection as a methodological criterion for recognizing entities such as social groups. For example, in a study of factors affecting the reproductive cycles of women in college dormitories, McClintock (1971, and in conversation) found that the initially randomly timed and different-length cycles of 135 women after several months became synchronized into 17 groups, each oscillating synchronously, in phase and with a common period. The members of these groups turned out to be those who spent most time together, as determined by sociological methods. After the onset of synchrony, group membership of an individual could be determined either from information about her reproductive cycle or from a sociogram representing her frequency of social interaction with other individuals. These groups are thus multiply detectable. This illustrates the point that there is nothing sacred about using perceptual criteria in individuating entities. The products of any scientific detection procedure, including procedures drawn from different sciences, can do as well, as Campbell suggests: "In the diagnosis of middle-sized physical entities, the boundaries of the entity are multiply confirmed, with many if not all of the diagnostic procedures confirming each other. For the more 'real' entities, the number of possible ways of confirming the boundaries is probably unlimited, and the more our knowledge expands, the more diagnostic means we have available. 'Illusions' occur when confirmation is attempted and found lacking, when boundaries diagnosed by one means fail to show up by other expected checks" (1958, pp. 23–24).

Illusions can arise in connection with robustness in a variety of ways. Campbell's remark points to one important way: Where expectations are derived from one boundary, or even more, the coincidence of several boundaries leads us to predict, assume, or expect that other relevant individuating boundaries will coincide. Perhaps most common, given the reductionism common today, are situations in which the relevant system boundary is in fact far more inclusive than one is led to expect from the coincidence of a number of boundaries individuating an object at a lower level. Such functional localization fallacies are found in neurophysiology, in genetics, in evolutionary biology (with the hegemony of the selfish gene at the expense of the individual or the group; see Wimsatt 1980b), in psychology, and (where it is a fallacy) with methodological individualism in the social sciences. In all these cases the primary object of analysis—be it a gene, a neuron, a neural tract, or an individual—may well be robust, but its high degree of entitativity leads us to hang too many boundaries and explanations on it. Where this focal entity is at a lower level, reductionism and robustness conspire to lead us to regard the higher-level systems as epiphenomenal. Another kind of illusion—the illusion that an entity is robust—can occur when the various means of detection supposed to be independent are not in fact. (This will be discussed further in the final section of this chapter.) Another kind of illusion or paradox arises particularly for functionally organized systems. This illusion occurs when a system has robust boundaries, but the different criteria used to decompose it into parts produce radically different boundaries. When the parts have little entitativity compared to the system, the holist's war cry (that the whole is more than the sum of the parts) will have a greater appeal. Elsewhere (Wimsatt 1974), I have explored this kind of case and its consequences for the temptation of antireductionism, holism, or, in extreme cases, vitalisms or ontological dualisms.

Robustness is a criterion for the reality of entities, but it also has played and can play an important role in the analysis of properties. Interestingly, the distinction between primary and secondary qualities, which had a central role in the philosophy of Galileo, Descartes, and Locke, can be made in terms of robustness. Primary qualities—such as shape, figure, and size—are detectable in more than one sensory modality. Secondary qualities—such as color, taste, and sound—are detectable through only one sense. I think it is no accident that seventeenth-century philosophers chose to regard primary qualities as the only things that were "out there"—in objects; their cross-modal detectability seemed to rule out their being products of sensory interaction with the world. By contrast the limitation of the secondary qualities to a single sensory modality seemed naturally to suggest that they were "in us," or subjective. Whatever the merits of the further seventeenth-century view that the secondary qualities were to be explained in terms of the interaction of a perceiver with a world of objects with primary qualities, this explanation represents an instance of an explanatory principle which is widely found in science (though seldom if ever explicitly recognized): *the explanation of that which is not robust in terms of that which is robust*. (For other examples see Wimsatt 1976a, pp. 243–249; Feynman 1965).

Paralleling the way in which Levins' use of robustness differs from Feynman's, *robustness, or the lack of it, has also been used in contexts where we are unsure about the status of purported properties, to argue for their veridicality or artifactuality*, and thus to discover the properties in terms of which we should construct our theories. This is the proposal of the now classic and widely used methodological paper of Campbell and Fiske (1959). Their convergent validity is a form of robustness, and their criterion of discriminant validity can be regarded as an attempt to guarantee that the invariance across test methods and traits is not due to their insensitivity to the variables under study. Thus, method bias, a common cause of failures of discriminant validity, is a kind of failure of the requirement for robustness that the different means of detection used are actually independent, in this case because the method they share is the origin of the correlations among traits.

Campbell and Fiske point out that very few theoretical constructs (proposed theoretical properties or entities) in the social sciences have significant degrees of convergent and discriminant validity, and they argue that this is a major difference between the social and natural or biological sciences—a difference which generates many of the problems of the social sciences. (For a series of essays which in effect claim that personality variables are highly context dependent and thus have very little or no robustness, see Shweder 1979a, b, 1980.)

While the natural and biological sciences have many problems where similar complaints could be made (the importance of interaction effects and context dependence is a key indicator of such problems), scientists in these areas have been fortunate in having at least a large number of cases where the systems, objects, and properties they study can be effectively isolated and localized, so that interactions and contexts can be ignored.

2.6 Robustness and Levels of Organization

Because of their multiplicity of connections and applicable descriptions, robust properties or entities tend to be (1) more easily detectable, (2) less subject to illusion or artifact, (3) more ex*planatorily fruitful*, and (4) *predictively richer than nonrobust properties or entities.* With this set of properties, it should be small wonder that we use robustness as a criterion for reality. It should also not be surprising that—since we view perception (as evolutionary epistemologists do) as an efficient tool for gathering information about the world—robustness should figure centrally in our analysis of perceptual hypotheses and heuristics (in Section 2.4 and in Section 2.7). Finally, since ready detectability, relative insensitivity to illusion or artifact, and explanatory and predictive fruitfulness are desirable properties for the components of scientific theories, we should not be surprised to discover that robustness is important in the discovery and description of phenomena (again, see Section 2.4) and in analyzing the structure of scientific theories (see Section 2.2).

One of the most ubiquitous phenomena of nature is its tendency to come in levels. If the aim of science, to follow Plato, is to cut up nature at its joints, then these levels of organization must be its major vertebrae. They have become so major, indeed, that our theories tend to follow these levels, and the language of our theories comes in strata. This has led many linguistically inclined philosophers to forgo talk of nature at all, and to formulate problems—for example, problems of reduction—in terms of "analyzing the relation between theoretical vocabularies at different levels." But our language, as Campbell (1974) would argue, is just another (albeit very important) tool in our struggle to analyze and to adapt to nature. In an earlier paper (Wimsatt 1976a, part III), I applied Campbell's criteria for entification to argue that entities at different levels of organization tend to be multiply connected in terms of their causal relations, primarily with other entities at their own level, and that they, and the levels they comprise, are highly robust. As a result, there are good explanatory reasons for treating different levels of organization as dynamically, ontologically, and epistemologically autonomous. There is no conflict here with the aims of good reductionistic science: there is a great deal to be learned about upper-level phenomena at lower levels of organization, but upper-level entities are not "analyzed away" in the process, because they remain robustly connected with other upper level entities, and their behavior is explained by upper-level variables.

To see how this is so, we need another concept—that of the *sufficient parameter,* introduced by Levins (1966, pp. 428 and 429):

> It is an essential ingredient in the concept of levels of phenomena that there exists a set of what, by analogy with the sufficient statistic, we can call sufficient parameters defined on a given level... which are very much fewer than the number of parameters on the lower level and which among them contain most of the important information about events on that level.
>
> The sufficient parameters may arise from the combination of results of more limited studies. In our robust theorem on niche breadth we found that temporal variation, patchiness of the environment, productivity of the habitat, and mode of hunting could all have similar effects and that they did this by way of their contribution to the uncertainty of the environment. Thus uncertainty emerges as a sufficient parameter.
>
> The sufficient parameter is a many-to-one transformation of lower-level phenomena. Therein lies its power and utility, but also a new source of imprecision. The many-to-one nature of "uncertainty" prevents us from going backwards. If either temporal variation or patchiness or low productivity leads to uncertainty, the consequences of uncertainty alone cannot tell us whether the environment is variable, or patchy, or unproductive. Therefore, we have lost information.

A sufficient parameter is thus a parameter, a variable, or an index which, either for most purposes or merely for the purposes at hand, captures the effect of significant variations in lower-level or less abstract variables (usually only for certain ranges of the values of these variables) and can thus he substituted for them in the attempt to build simpler models of the upper-level phenomena.

Levins claims that this notion is a natural consequence of the concept of levels of phenomena, and this is so, though it may relate to degree of abstraction as well as to degree of aggregation. (The argument I will give here applies only to levels generated by aggregation of lower-level entities to form upper-level ones.) Upper-level variables, which give a more "coarse-grained" description of the system, are much smaller in number than the lower-level variables necessary to describe the same system. Thus, there must be, for any given degree of resolution between distinguishable state descriptions, far fewer distinguishable upper-level state descriptions than lower-level ones. The smaller number of distinguishable upper-level states entails that for any given degree of resolution, there must be many-one mappings between at least some lower-level and upper-level state descriptions with many lower-level descriptions corresponding to a single upper-level description. But then, those upper-level state descriptions with multiple lower-level state descriptions are robust over changes from one of these lower-level descriptions to another in its set.

Furthermore, the stability of (and possibility of continuous change in) upper-level phenomena (remaining in the same macrostate or changing by moving to neighboring states) places constraints on the possible mappings between lower-level and upper-level states: in the vast majority of cases neighboring microstates must map without discontinuity into the same or neighboring macrostates; and, indeed, most local microstate changes will have no detectable macrolevel effects. *This fact gives upper-level phenomena and laws a certain insulation from* (through their invariance over: robustness again!) *lower-level changes and generates a kind of explanatory and dynamic (causal) autonomy of the upper-level phenomena and processes,* which I have argued for elsewhere (Wimsatt 1976a, pp. 249–251, 1976b).

If one takes the view that causation is to be characterized in terms of manipulability (see, for example, Gasking 1955; Cook and Campbell 1979), the fact that the vast majority of manipulations at the microlevel do not make a difference at the macrolevel means that macrolevel variables are almost always more causally efficacious in making macrolevel changes than microlevel variables. This gives explanatory and dynamic autonomy of the upper-level entities, phenomena, laws, and relations, within a view of explanation which is sensitive to problems of computational complexity and the costs and benefits we face in offering explanations. As a result, it comes much closer than the traditional hypotheticodeductive view to being able to account for whether we explain a phenomenon at one level and when we choose to go instead to a higher or lower level for its explanation (see Wimsatt 1976a, part III, and 1976b, particularly sections 4, 5, 6, and the appendix.)

The many-one mappings between lower- and upper-level state descriptions mentioned above are consistent with correspondences between types of entities at lower and upper levels but do not entail them. There may be only token-token mappings (piece-meal mappings between instances of concepts, without any general mappings between concepts), resulting in the upper-level properties being supervenient on rather than reducible to lower-level properties (Kim 1978; Rosenberg 1978). The main difference between Levins' notion of a sufficient parameter and the notion of supervenience is that the characterization of supervenience is embedded in an assumed apocalyptically complete and correct description of the lower and upper levels. Levins makes no such assumption and defines the sufficient parameter in terms of the imperfect and incomplete knowledge that we actually have of the systems we study. It is a broader and less demanding notion, involving a relation which is inexact, approximate, and admits of both unsystematic exceptions (requiring a *ceteris paribus* qualifier) and systematic ones (which render the relationship conditional).

Supervenience could be important for an omniscient Laplacean demon but not for real, fallible, and limited scientists. The notion of supervenience could be regarded as a kind of ideal limiting case of a sufficient parameter as we come to know more and more about the system, but it is one which is seldom if ever found in the models of science. The concept of a sufficient parameter, by contrast, has many instances in science. It is central to the analysis of reductive explanation (Wimsatt 1976a, b, pp. 685–689, 1979) and has other uses as well (Wimsatt 1980a, section 4).

2.7 Heuristics and Robustness

Much or even most of the work in philosophy of science today which is not closely tied to specific historical or current scientific case studies embodies a metaphysical stance which, in effect, assumes that the scientist is an omniscient and computationally omnipotent Laplacean demon. Thus, for example, discussions of reductionism are full of talk of "in principle analyzability" or "in principle deducibility," where the force of the "in principle" claim is held to be something like "If we knew a total description of the system at the lower level, and all the lower-level laws, a sufficiently complex computer could generate the analysis of all the upper-level terms and laws and predict any upper-level phenomenon." Parallel kinds of assumptions of omniscience and computational omnipotence are found in rational decision theory, discussions of Bayesian epistemology, automata theory and algorithmic procedures in linguistics and the philosophy of mind, and the reductionist and foundationalist views of virtually all the major figures of twentieth-century logical empiricism. It seems almost to be a corollary to a deductivist approach to problems in philosophy of science (see Wimsatt 1979) and probably derives ultimately from the Cartesian vision criticized earlier in this chapter.

I have already written at some length attacking this view and its application to the problem of reduction in science (see Wimsatt 1974, 1976a, pp. 219–237, 1976b, 1979, 1980b, section 3; and also Boyd 1972). The gist of this attack is threefold: (1) On the "Laplacean demon" interpretation of "in principle" claims, we have no way of evaluating their warrant, at least in science. (This is to be distinguished from cases in mathematics or automata theory, where "in principle" claims can be explicated in terms of the notion of an effective procedure.) (2) We are in any case not Laplacean demons, and a philosophy of science which could have normative force only for Laplacean demons thus gives those of us who do not meet these demanding specifications only counterfactual guidance. That is, it is of no real use to practicing scientists and, more strongly, suggests methods and viewpoints which are less advantageous than those derived from a more realistic view of the scientist as problem solver (see Wimsatt 1979). (3) An alternative approach, which assumes more modest capacities of practicing scientists, does provide real guidance, better fits with actual scientific practice, and even (for reductive explanations) provides a plausible and attractive alternative interpretation for the "in principle" talk which so many philosophers and scientists use frequently (see Wimsatt 1976a, part II; 1976b, pp. 697–701).

An essential and pervasive feature of this more modest alternative view is the replacement of the vision of an ideal scientist as a computationally omnipotent algorithmizer with one in which the scientist as decision maker, while still highly idealized, must consider the size of computations and the cost of data collection, and in other very general ways must be subject to considerations of efficiency, practical efficacy, and cost-benefit constraints. This picture has been elaborated over the last twenty-five years by Herbert Simon and his coworkers, and their ideal is "satisficing man," whose rationality is bounded, by contrast with the unbounded omniscience and computational omnipotence of the "economic man" of rational decision theory (see Simon 1955, reprinted as chapter 1 of Simon 1979; see also Simon 1969, 1973). Campbell's brand of fallibilism and critical realism from an evolutionary perspective also place him squarely in this tradition.

A key feature of this picture of man as a boundedly rational decision maker is the use of heuristic principles where no algorithms exist or where the algorithms that do exist require an excessive amount of information, computational power, or time. I take a heuristic procedure to have three important properties (see also Wimsatt 1980b, section 3): (1) By contrast with an algorithmic procedure (here ignoring probabilistic automata), *the correct application of a heuristic procedure does not guarantee a solution* and, if it produces a solution, does not guarantee that the solution is correct. (2) *The expected time, effort, and computational complexity of producing a solution with a heuristic procedure is appreciably less* (often by many orders of magnitude for a complex problem) *than that expected with an algorithmic procedure.* This is indeed the reason why heuristics are used. They are a cost-effective way, and often the *only* physically possible way, of producing a solution. (3) *The failures and errors produced when a heuristic is used are not random but systematic. I* conjecture that *any heuristic, once we understand how it works, can be made to fail.* That is, given this knowledge of the heuristic procedure, we can construct classes of problems for which it will always fail to produce an answer or for which it will always produce the wrong answer. This property of systematic production of wrong answers will be called the *bias* of the heuristic.

This last feature is exceedingly important. Not only can we work forward from an understanding of a heuristic to predict its biases, but we can also work backward from the observation of systematic biases as data to hypothesize the heuristics which produced them; and if we can get independent evidence (for example, from cognitive psychology) concerning the nature of the heuristics, we can propose a well-founded explanatory and predictive theory of the structure of our reasoning in these areas. This approach was implicitly (and sometimes explicitly) followed by Tversky and Kahneman (1974), in their analysis of fallacies of probabilistic reasoning and of the heuristics which generate them (see also Shweder 1977, 1979a, b, 1980, for further applications of their work; and Mynatt, Doherty, and Tweney 1977, for a further provocative study of bias in scientific reasoning). The systematic character of these biases also allows for the possibility of modifications in the heuristic or in its use to correct for them (see Wimsatt 1980b, pp. 52–54).

The notion of a heuristic has far greater implications than can be explored in this chapter. In addition to its centrality in human problem solving, it is a pivotal concept in evolutionary biology and in evolutionary epistemology. It is a central concept in evolutionary biology because any biological adaptation meets the conditions given for a heuristic procedure. First, it is a commonplace among evolutionary biologists that adaptations, even when functioning properly, do not guarantee survival and production of offspring. Second, they are, however, cost-effective ways of contributing to this end. Finally, any adaptation has systematically specifiable conditions, derivable through an understanding of the adaptation, under which its employment will actually decrease the fitness of the organism employing it, by causing the organism to do what is, under those conditions, the wrong thing for its survival and reproduction. (This, of course, seldom happens in the organism's normal environment, or the adaptation would become maladaptive and be selected against.) This fact is indeed systematically exploited in the functional analysis of organic adaptations. It is a truism of functional inference that learning the conditions under which a system malfunctions, and how it malfunctions under those conditions, is a powerful tool for determining how it functions normally and the conditions under which it was designed to function. (For illuminating discussions of the problems, techniques, and fallacies of functional inference under a variety of circumstances, see Gregory 1958; Lorenz 1965; Valenstein 1973; Glassman 1978.)

The notion of a heuristic is central to evolutionary epistemology because Campbell's (1974, 1977) notion of a vicarious selector, which is basic to his conception of a hierarchy of adaptive and selective processes spanning subcognitive, cognitive, and social levels, is that of a heuristic procedure. For Campbell a vicarious selector is a substitute and less costly selection procedure acting to optimize some index which is only contingently connected with the index optimized by the selection process it is substituting for. This contingent connection allows for the possibility—indeed, the inevitability—of systematic error when the conditions for the contingent concilience of the substitute and primary indices are not met. An important ramification of Campbell's idea of a vicarious selector is the possibility that one heuristic may substitute for another (rather than for an algorithmic procedure) under restricted sets of conditions, and that this process may be repeated, producing a nested hierarchy of heuristics. He makes ample use of this hierarchy in analyzing our knowing processes (Campbell 1974). I believe that this is an appropriate model for describing the nested or sequential structure of many approximation techniques, limiting operations, and the families of progressively more realistic models found widely in progressive research programs, as exemplified in the development of nineteenth-century kinetic theory, early twentieth-century genetics, and several areas of modern population genetics and evolutionary ecology.

To my mind, Simon's work and that of Tversky and Kahneman have opened up a whole new set of questions and areas of investigation of pragmatic inference (and its informal fallacies) in science, which could revolutionize our discipline in the next decade. (For a partial view of how studies of reduction and reductionism in science could be changed, see Wimsatt 1979.) This change in perspective would bring philosophy of science much closer to actual scientific practice without surrendering a normative role to an all-embracing descriptivism. And it would reestablish ties with psychology through the study of the character, limits, and biases of processes of empirical reasoning. Inductive procedures in science are heuristics (Shimony 1970), as are Mill's methods and other methods for discovering causal relations, building models, and generating and modifying hypotheses.

Heuristics are also important in the present context, because the procedures for determining robustness and for making further application of these determinations for other ends are all heuristic procedures. Robustness analysis covers a class of powerful and important techniques, but they are not immune to failures. There are no magic bullets in science, and these are no exception.

Most striking of the ways of failure of robustness analysis is one which produces illusions of robustness: the failure of the different supposedly independent tests, means of detection, models, or derivations to be truly independent. This is the basis for a powerful criticism of the validity of IQ scales as significant measures of intelligence (see McClelland 1973). Failures of independence are not easy to detect and often require substantial further analysis. Without that, such failures can go undetected by the best investigators for substantial lengths of time. Finally, the fact that different heuristics can be mutually reinforcing, each helping to hide the biases of the others (see Wimsatt 1980b, sections 5 and 8), can make it much harder to detect errors which would otherwise lead to discovery of failures of independence. The failure of independence in its various modes, and the factors affecting its discovery, emerges as one of the most critical and important problems in the study of robustness analysis, as is indicated by the history of the group selection controversy.

2.8 Robustness, Independence, and Pseudorobustness: A Case Study

In recent evolutionary biology (since Williams' seminal work in 1966), group selection has been the subject of widespread attack and general suspicion. Most of the major theorists (including W.D. Hamilton, John Maynard Smith, and E.O. Wilson) have argued against its efficacy. A number of mathematical models of this phenomenon have been constructed, and virtually all of them (see Wade 1978) seem to support this skepticism. The various mathematical models of group selection surveyed by Wade all admit of the possibility of group selection. But almost all of them predict that group selection should only very rarely be a significant evolutionary factor; that is, they predict that group selection should have significant effects only under very special circumstances—for extreme values of parameters of the models—which should seldom be found in nature. Wade undertook an experimental test of the relative efficacy of individual and group selection—acting in concert or in opposition in laboratory populations of the flour beetle, *Tribolium.* This work produced surprising results. Group selection appeared to be a significant force in these experiments, one capable of overwhelming individual selection in the opposite direction for a wide range of parameter values. This finding, apparently contradicting the results of all of the then extant mathematical models of group selection, led Wade (1978) to a closer analysis of these models, with results described here.

All the models surveyed made simplifying assumptions, most of them different. Five assumptions, however, were widely held in common; of the twelve models surveyed, each made at least three of these assumptions, and five of the models made all five assumptions. Crucially, for present purposes, the five assumptions are biologically unrealistic and incorrect, and each independently has a strong negative effect on the possibility or efficacy of group selection. It is important to note that these models were advanced by a variety of different biologists, some sympathetic to and some skeptical of group selection as a significant evolutionary force. Why, then, did all of them make assumptions strongly inimical to it? Such a coincidence, radically improbable at best, cries out for explanation: we have found a systematic bias suggesting the use of a heuristic.

These assumptions are analyzed more fully elsewhere (Wade 1978; Wimsatt 1980b). My discussion here merely summarizes the results of my earlier analysis, where (in Section 2.5) I presented a list of nine reductionistic research and modeling strategies. Each is a heuristic in that it has systematic biases associated with it, and these biases will lead to the wrong answer if the heuristic is used to analyze certain kinds of systems. It is the use of these heuristics, together with certain "perceptual" biases (deriving from thinking of groups as "collections of individuals" rather than as robust entities analogous to organisms), that is responsible for the widespread acceptance of these assumptions and the almost total failure to notice what an unrealistic view they give of group selection. Most of the reductionistic heuristics lead to a dangerous oversimplification of the environment being studied and a dangerous underassessment of the effects of these simplifications. In the context of the perceptual bias of regarding groups as collections of individuals (or sometimes even of genes), the models tend systematically to err in the internal and relational structure they posit for the groups and in the character of processes of group reproduction and selection.

The first assumption, that the processes can be analyzed in terms of selection of alternative alleles at a single locus, is shown to be empirically false by Wade's own experiments, which show conclusively that both individual and group selection is proceeding on multilocus traits. (For an analysis of the consequences of treating a multilocus trait erroneously as a single-locus trait, see Wimsatt 1980b, section 4) The fifth assumption, that individual and group selections are opposed in their effects, also becomes untenable for a multilocus trait (see Wimsatt 1980b, section 7).

The second assumption is equivalent to the time-honored assumption of panmixia, or random mating within a population, but in the context of a group selection model it is equivalent to assuming a particularly strong form of blending inheritance for group inheritance processes. This assumption is factually incorrect and, as Fisher showed in 1930, effectively renders evolution at that level impossible. The third assumption is equivalent to assuming that groups differ in their longevity but not in their reproductive rates. But, as all evolutionary biologists since Darwin have been aware, variance in reproductive rate has a far greater effect on the intensity of selection than variance in longevity. So the more significant component was left out in favor of modeling the less significant one. (The second and third assumptions are discussed in Wimsatt 1980b, section 7.) The fourth assumption is further discussed and shown to be incorrect in Wade (1978).

The net effect is a set of cumulatively biased and incorrect assumptions, which, not surprisingly, lead to the incorrect conclusion that group selection is not a significant evolutionary force. If I am correct in arguing that these assumptions probably went unnoticed because of the biases of our reductionistic research heuristics, a striking analogy emerges. The phenomenon appeared to be a paradigmatic example of Levinsian robustness. A wide variety of different models, making different assumptions, appeared to show that group selection could not be efficacious. But the robustness was illusory, because the models were not independent in their assumptions. The commonality of these assumptions appears to be a species of method bias, resulting in a failure of discriminant validity (Campbell and Fiske 1959). But the method under consideration is not the normal sort of test instrument that social scientists deal with. Instances of the method are reductionistic research heuristics, and the method is reductionism. For the purposes of problem solving, our minds can be seen as a collection of methods, and the particularly single-minded are unusually prone to method bias in their thought processes. This conclusion is ultimately just another confirmation at another level of something Campbell has been trying to teach us for years about the importance of multiple independent perspectives.

References

Allen, G.E. 1979. *Thomas Hunt Morgan: The Man and his Science*. Princeton, NJ: Princeton University Press.

Barlow, R.E., and F. Proschan. 1975. *The Mathematical Theory of Reliability and Life Testing*. New York: Wiley.

Boyd, R. 1972. "Determinism, Laws, and Predictability in Principle." *Philosophy of Science* 39:431–50.

Campbell, D.T. 1958. "Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities." *Behavioral Science* 3:14–25.

Campbell, D.T. 1966. "Pattern Matching as an Essential in Distal Knowing." In *The Psychology of Egon Brunswik*, edited by K.R. Hammond, 81–106. New York: Holt, Rinehart and Winston.

Campbell, D.T. 1969a. "Definitional Versus Multiple Operationalism." Reprinted in *Methodology and Epistemology for Social Science*, edited by E.S. Overman. Chicago, IL: University of Chicago Press, 1988.

Campbell, D.T. 1969b. "Prospective: Artifact and Control." In *Artifact in Behavioral Research*, edited by R. Rosenthal and R. Rosnow, 351–82. New York: Academic.

Campbell, D.T. 1974. "Evolutionary Epistemology." In *The Philosophy of Karl Popper*, edited by P.A. Schilpp, 412–63. La Salle, IL: Open Court.

Campbell, D.T. 1977. *Descriptive Epistemology: Psychological, Sociological, and Evolutionary*. William James Lectures, Harvard University. Reprinted in *Methodology and Epistemology for Social Science*, edited by E.S. Overman. Chicago, IL: University of Chicago Press, 1988.

Campbell, D.T., and D.W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait–Multimethod Matrix." *Psychological Bulletin* 56:81–105.

Cook, T.D., and D.T. Campbell. 1979. *Quasi Experimentation: Design and Analysis for Field Settings*. Chicago: Rand McNally.

Cronbach, L.J., and P.E. Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52:281–302.

Darden, L. 1974. "Reasoning in Scientific Change: The Field of Genetics at its Beginnings." Unpublished Doctoral diss., Committee on the Conceptual Foundations of Science, University of Chicago.

Feynman, R.P. 1965. *The Character of Physical Law*. Cambridge, MA: MIT Press.

Fisher, R.A. 1930. *The Genetic Theory of Natural Selection*. New York: Clarendon Press.

Gasking, D.A.T. 1955. "Causation and Recipes." *Mind* 64:479–87.

Glassman, R.B. 1978. "The Logic of the Lesion Experiment and its Role in the Neural Sciences." In *Recovery from Brain Damage: Research and Theory*, edited by S. Finger, 3–31. New York: Plenum.

Glymour, C. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Gregory, R.L. (l958) "Models and the localization of function in the central nervous system." In National Physical Laboratory, Symposium No. 10. Mechanization of Thought Processes, vol. 2, H.M.S., reprinted in his (1974) Concepts and Mechanisms of Perception, Duckworth, 566–83.

Kauffman, S.A. 1971. "Gene Regulation Networks: A Theory of their Structure and Behavior." In *Current Topics in Developmental Biology*, edited by A. Moscona and A. Monroy, vol. 6, 145–82. New York: Academic Press.

Kim, J. 1978. "Supervenience and Nomological Incommensurables." *American Philosophical Quarterly* 15:149–56.

Laudan, L. 1971. "William Whewell on the Consilience of Inductions." *The Monist* 55:368–91.

Levins, R. 1966. "The Strategy of Model Building in Population Biology". *American Scientist* 54:421–31.

Levins, R. 1968. *Evolution in Changing Environments*. Princeton, NJ: Princeton University Press.

Lorenz, K.Z. 1965. *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.

Margenau, H. 1950. *The Nature of Physical Reality*. New York: McGrawHill.

McClelland, D.D. 1973. "Testing for Competence Rather Than for Intelligence." *American Psychologist* 29:107.

McClintock, M.K. 1971. "Menstrual Synchrony and Suppression." *Nature* 229:244–5.

Moore, J.A. 1972. *Heredity and Development*. 2nd ed. New York: Oxford University Press.

Mynatt, C.R., M.E. Doherty, and R.D. Tweney. 1977. "Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference." *Quarterly Journal of Experimental Psychology* 29:85–95.

Olby, R. 1974. *The Path to the Double Helix*. Seattle, WA: University of Washington Press.

Omanson, R.C. 1980a. *The Narrative Analysis: Identifying Central, Supportive and Distracting Content*. Unpublished manuscript.

Omanson, R.C. 1980b. "The Effects of Centrality on Story Category Saliency: Evidence for Dual Processing." Paper presented at 88th annual meeting of the American Psychological Association, Montreal, September 1980.

Peirce, C.S. 1936. "Some Consequences of Four Incapacities." In *Collected Papers of Charles Sanders Peirce*, edited by C. Hartshorne and P. Weiss, vol. 5. Cambridge, MA: Harvard University Press. (Originally published 1868.)

Putnam, H. 1962. "The Analytic and the Synthetic." In *Minnesota Studies in the Philosophy of Science*, edited by H. Feigl and G. Maxwell, vol. 3, 358–97. Minneapolis: University of Minnesota Press.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Rosenberg, A. 1978. "The Supervenience of Biological Concepts." *Philosophy of Science* 45:368–86.

Shimony, A. 1970. "Statistical Inference." In *The Nature and Function of Scientific Theories*, edited by R.G. Colodny, 79–172. Pittsburgh, PA: University of Pittsburgh Press.

Shweder, R.A. 1977. "Likeness and Likelihood in Everyday Thought: Magical Thinking in Judgements About Personality." *Current Anthropology* 18:637–48; reply to discussion, 652–8.

Shweder, R.A. 1979a. "Rethinking Culture and Personality Theory. Part I." *Ethos* 7:255–78.

Shweder, R.A. 1979b. "Rethinking Culture and Personality Theory. Part II." *Ethos* 7:279–311.

Shweder, R.A. 1980. "Rethinking Culture and Personality Theory. Part III." *Ethos* 8:60–94.

Simon, H. (1955). "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69:99–118.

Simon, H.A. 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Simon, H.A. 1973. "The Structure of Ill–Structured Problems." *Artificial Intelligence* 4:181–201.

Simon, H.A. 1979. *Models of Thought*. New Haven, CT: Yale University Press.

Tversky, A., and D. Kahneman. 1974. "Decision Under Uncertainty: Heuristics and Biases." *Science* 185:1124–31.

Valenstein, E. 1973. *Brain Control*. New York: Wiley.

von Neumann, J. 1956. "Probabilistic Logic and the Synthesis of Reliable Organisms from Unreliable Components." In *Automata Studies*, edited by C.E. Shannon and J. McCarthy, 43–98. Princeton, NJ: Princeton University Press.

Wade, M.J. 1978. "A Critical Review of the Models of Group Selection." *Quarterly Review of Biology* 53(3):101–14.

Williams, G.C. 1966. *Adaptations and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.

Wimsatt, W.C. 1974. "Complexity and Organization." In *Proceedings of the Meetings of the Philosophy of Science Association, 1972*, edited by K.F. Schaffner and R.S. Cohen, 67–86. Dordrecht, Netherlands: Reidel.

Wimsatt, W.C. 1976a. “Reductionism, Levels of Organization, and the Mind-Body Problem.” In *Consciousness and the Brain: Scientific and Philosophical Strategies*, edited by G.G. Globus, G. Maxwell, and I. Savodnik, 199–267. New York: Plenum.

Wimsatt, W.C. 1976b. “Reductive Explanation: A Functional Account.” In *Proceedings of the Meetings of the Philosophy of Science Association, 1974*, edited by C.A. Hooker, G. Pearse, A.C. Michalos, and J.W. van Evra, 671–710. Dordrecht, Netherlands: Reidel.

Wimsatt, W.C. 1979. “Reduction and Reductionism.” In *Current Problems in Philosophy of Science*, edited by P.D. Asquith and H. Kyburg, Jr., 352–77. East Lansing, MI: Philosophy of Science Association.

Wimsatt, W.C. 1980a. “Randomness and Perceived Randomness in Evolutionary Biology.” *Synthese* 43(3):287–329.

Wimsatt, W.C. 1980b. “Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy.” In *Scientific Discovery*, edited by T. Nickles, 213–59. Historical and scientific case studies, vol. 2. Dordrecht, Netherlands: Reidel.

Chapter 3
Robustness: Material, and Inferential, in the Natural and Human Sciences

William C. Wimsatt

3.1 Robustness Introduced: Historical Background and Stage Setting

When I came to Chicago as a post-doc in the summer of 1969, to work with Richard Lewontin, I also met Richard Levins, a deeply reflective, politically active, and strikingly creative mathematical ecologist. Both were to have a deep influence on how I saw science. Levins was the author of a remarkable paper that I had read as a graduate student, published in *American Scientist* in 1966, "The Strategy of Model Building in Population Biology". Among several other innovative ideas, sketched there and then further elaborated in his 1968 book, *Evolution in Changing Environments*, Levins proposed a methodology of looking for "robust theorems."

Levins' title was immediately arresting. Philosophers then didn't talk about (heuristic) strategies, model-building or population biology. This was totally virgin territory. Model-building was a new topic for both biologists and philosophers. Both talked about theory—and treated their equations as completed edifaces.[1] For philosophers, the only relevant contrast was between observation (or empirical evidence) that was given from nature and trusted, and theory—which was constructed by humans, and the best that we had, but still suspect. Theories were confirmed, disconfirmed, or yet to be tested. The context of justification gave the only subject matter then deemed appropriate for philosophers. Discovery or problem-solving was

[1] The first textbook in this area that explicitly recognized that the equations presented were models and not established theories, and self-consciously discussed their shortcomings and idealizations, was Wilson and Bossert's (1971) *A Primer of Population Biology*. An inexpensive paperback, designed as a supplement to "main" biology texts, it was also the first book published by Andy Sinauer's new firm which targeted and became the premier publisher in this area.

W.C. Wimsatt (✉)
Department of Philosophy and Conceptual and Historical Studies of Science,
The University of Chicago, Chicago, IL, USA

Center for Philosophy of Science, University of Minnesota, St. Paul, MN, USA
e-mail: wwim@uchicago.edu

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_3,

held to be idiosyncratic and unsystematic. Models (insofar as they were spoken of by philosophers at all) were interpreted instanciations of theory—in a formal-logic driven account that gave us "the semantic conception of theories"—an idea that as far as I can tell has borne no fruit outside of philosophy. The ideas that there might be different kinds of theory or levels of theoretical activity, that theory was full of (non-deductive) heuristic approximations and articulations, and that model-building might be a structured and heuristic, but particularly tentative and exploratory activity with known false or oversimplified conceptual tools (Wimsatt 1987) were all beyond the pale. And that investigators tried to model phenomena (rather than theories) was thought to be conceptual incoherence (a view forcefully pushed by both Donald Davidson and Patrick Suppes).[2] From my experience at the time, topics connected with problem-solving and discovery were thought to be inappropriate or at best marginal for philosophers in any case.

But here were Lewontin, who had written a thought provoking paper on modeling in 1963 discussing his pioneering computer simulations of group selection (the t-allele in the house mouse, *Mus musculus*) and another particular 2-locus evolutionary problem (with fitness interactions between chromosome inversions in the Australian grasshopper, *Moraba scurra*) and Levins for whom theoretically informed (but more loosely connected) mathematical structures could be qualitatively analyzed without solving them—even if one had only partial information about the system (Levins 1974). For both of them, modeling was the cutting edge in the investigation of nature. They were the best in the business, and their work was fascinating. The philosophers had to be wrong.

Population biology was also a new subject—crafted by a number of innovative biologists—then young turks, but now almost legendary for the theoretical, empirical, and conceptual innovations they have spawned in the last 50 years. The group included Levins, Lewontin, Robert MacArthur, Ed Wilson, Egbert Leigh, and Leigh van Valen. The new perspective was crafted when they spent summers together in Vermont in the 1960s. They had concluded that ecology and the population genetics dominated evolutionary theory of the New Synthesis needed to be more closely articulated. These two fields, though in principle connected, had developed with at best weak links between them. There were no inter-level derivational or deductive relations to unify them. Community ecology had developed with conceptual tools derived for modeling the interactions of populations of diverse species who were born, lived, and died, (perhaps with multiple age-classes with age-specific birth and death rates), and who predated, parasitized, cooperated, and competed with one another in ways symbolized by one-dimensional lumped interaction coefficients. (Levins' diagram of the structure of population biology was more complex—it contained 25 boxes of local models connected in a directed graph impinging on

[2] I was surprised at the vehemence they showed in defending this belief. I remember describing scientific practice to Davidson, and saying that biologists must have meant something different by models than he did. His response was simply to say angrily: "Well, they're wrong!" (end of conversation). Such hubris! This strange excursion from the real world is no-where better described and criticized than in Downes (1992).

community structure—the core model of ecology). To look just at how the simplifications in this area were related to population genetics, they didn't: in these models members of the same species were treated as genetically and behaviorally identical, and genetic variation was supposed to be irrelevant because any evolution was supposed to take place over much longer time spans.

Population genetics had correspondingly developed to emphasize only (a minimalist description of) the genetic structure of individuals, and simple models of who mated with whom. Fitnesses were add-ons to the models—generally postulated or measured in laboratory experiments and (very occasionally) in nature. But basically the phenotype was just treated as a black-box scalar multiplier for gene frequencies with (often arbitrarily assigned) selection coefficients. There were genes and chromosomes (or rather linkage coefficients) but nothing more complex than genes at 2 loci, and most theory and applications were pursued with single-locus models. Changes were assumed to take place over many generations, far longer than the seasonal fluctuations of ecological populations. There was no development and no physiology in such models, no ecology, no temporal variation of the environment, and a population structure not realistically hooked into the real spatial and temporal heterogeneity of the environment. So though basically of the same processes on different time scales, the relevant rates of change were held to be so different that they could be treated as decoupled. Both areas were grossly oversimplified.

The conviction that the articulation had to be made grew out of several things. Levins, MacArthur and Wilson were convinced that multiple components of fitness that keyed into diverse ecological and organizational conditions could show interesting tradeoffs (reflected most strongly in Levins' fitness set analysis and Wilson's ergonomics of castes and tasks in social insects) which they wanted to articulate with population genetic models. The theory of island biogeography of MacArthur and Wilson put more qualitative substance into local colonizations and extinctions in patchy environments that gave evolution an ecological time scale and population structures that also put Sewall Wright's population genetic theories involving small populations and a metapopulation structure into natural focus. And Leigh Van Valen, a polydisciplinary paleontologist urged a view of evolution as "the control of development by ecology" (Van Valen 1989). The gel electrophoresis applied by Hubby and Lewontin to sample protein (and thus genetic) variation in natural populations in *Drosophila* was rapidly diffused to other studies that showed seasonal variations of gene frequencies in voles, and spawned a whole range of studies in a similar vein. More recently, three generations of work by Rosemary and Peter Grant and their many talented students integrated ecology, seasonality, and mating history in multiple species of Galapagos finches and their vegetable and insect prey as measured in rich genetic, demographic and ecological detail, producing perhaps the most integrated studies in population biology to date—though much of this occurred well after the founding conceptual changes that brought the new hybrid discipline into being (Grant 1999).

This situation generated two important stimuli. First, phenomena in two fields were shown to be significantly coupled to each other. And second, mathematical ecology, demography (for age structure of populations) and population genetics, then the three most mathematized fields in what were to become the Darwinian sciences, had a new need for comparability and articulation. But the complexity that had to be dealt with was potentially overwhelming. What to do? Many things, but the field was set for Levins' search for "robust theorems" that were true across multiple models that made different assumptions about a given phenomenon or relationship. In the face of incomplete data, with many practically unmeasurable key parameters, and the fear that many outcomes might be very sensitive to values of these variables or to as yet undetermined unrealistic descriptions in the models, the idea of an analytic comparison of models that might possibly span the range of possible states, relationships, and outcomes to look for robust results seemed both ingenious and attractive. But we needed to be at as much interested in models where the results broke down, since these showed the limits of robustness for the result. "Randomness and Perceived-Randomness in Evolutionary Biology" (Wimsatt 1980a) was in part a study of certain kinds of uncertainty (that Levins had used as an example of a "sufficient parameter" in search of a robust theorem) and also a study of and argument for the robustness of chaotic behavior in ecological systems.

I discussed Campbell's explicit writings developing what he called "triangulation" (e.g., Campbell 1966) more extensively in my 1981 paper on robustness, but the main distinguishing feature of his analysis was to emphasize the use of multiple perceptual methods to "triangulate" on the properties of objects not immediately known (things that were "distal"). He wrote far more on it than Levins, and had to deal with what was apparently a far more challenging context—namely the assessment of human characteristics using a variety of social science measures. For this he drew on analogies with perception—then a familiar area to psychologists. He also sought to argue for the "entitativity" of social collectives in terms of the agreement of multiple criteria for individuating them—or more generally on the reality of social factors that were insufficiently recognized by methodological individualists using the same criteria—multiple independent means of detection or measurement. His main challenge, I believe was the greater difficulty of distinguishing between parts of the detection instrument, or its effects and parts of the capability of object under study in the human sciences, although it is arguable that many artifacts in the natural science share exactly this feature. Indeed, I believe that this was just what characterized the "bacterial mesosome"—an artifact of preparation that was mistaken for a feature of cellular ultrastructure—discussed by Rasmussen (1993), Culp (1994, 1995), and Hudson (1999).

At the time I wrote the paper on robustness analysis for the Campbell festschrift, I tried to systematize the method, recognizing that robustness analysis proceeded from analysis of failures and limits of robustness as well as successes. I tried to draw together every methodological variation I could find that shared the use of multiple means of access, detection, inference or deduction to secure a more substantial handle on the phenomenon, object, process, or result under study, and found confirmations of it in some unexpected places. The resulting collection was very heterogeneous, but that was part of the point. All of these practices had some interesting things in common, despite their differences, and I wanted to mark these. But any comparative analysis must look to the differences as well as the similarities.

Others since (Weisberg 2006, Weisberg and Riesman 2008, Willinger and Doyle 2005, and Nickles, this volume) have noted the variety, and distinguished and elaborated some important sub-classes of robustness (Calcott 2010, and Raerinne 2010). I will make summary remarks on what are plausibly three main sub-types of robust inference, and then turn to a fuller commentary on the fourth, which is to look for and assess the significance of and use robustness in material systems.[3]

In my survey of eight kinds of applications in 1981, I specifically included three classes of cases:

(1) Multiple means of empirical access or interaction with the target of investigation, in which I would include both different sensory modalities, and different instrument mediated paths to convergent or complementary result(s). This is in accord with our normal intuitions of regarding detection instruments as extensions of our sensory systems. Some of these amplify or magnify our existing senses to different size, time, or sensitivity scales, while others exploit different sensory modalities (e.g., magnetic anomaly detection) or sensory capabilities beyond the frequency range of our senses (e.g., UV or infra-red, or ultrasonic transmissions.) I had been thinking of radio-telescopic and light telescopic observation of the same objects, but I also included the possibility of theory-mediated analysis of the signal, such as an appropriately time-lagged correlation required to pick orbiting Phobos out of successive Mars images rather than have it averaged out as noise). The lovely study by Allamel-Raffin and Gangloff (Chapter 7) documenting robustness analysis with multi-spectral images using different instruments with different frequency filters and spatial resolutions elaborates the richness of this kind of analysis. Their analysis also involved a kind of calibration and also showed the use of theoretical knowledge of what kinds of emissions should covary in their robustness arguments. My original analysis argued for the importance of (the at least partial) independence of the detection channels, and this analysis was carried significantly further through the first-hand expertise of the talented and prematurely deceased Culp (1994, 1995).

(2) A second important means is the use of different analytical methods. Here I included multiple derivational paths within a single multiply connected theory (Feynman 1967) and also multiple independently motivated and constructed models from which a common result could be derived (Levins 1966). Glymour (1980) clearly envisioned the possibility of the former kind of relationship

[3] Brett Calcott (2010) has since produced a nice three-way classification in terms of multiple means between formal methods, detection, and material robustness (lumping the 1st and 3rd senses of the 4 discussed here). I agree with his classification: the distinction between my 1 and 3 can be seen as a distinction between kinds of detection methods, though it is also true that a model (from the "false model" perspective) can be seen as a kind of selective filter for the kind of pattern explored in the model. With that, only the kinds of exploration of the functions of multiple derivational pathways in a single formal system (like those investigated by Kromer here and by Corfield 2010) would stand out as a distinct kind. But to regard models as kinds of selective pattern detectors might violate too many intuitions!

though his network connections are ambiguous between analytic and empirical links. Ralf Krömer's (Chapter 8) work here breaks new ground in the richness of the functions he found for robustness analysis in mathematics, and how it made sense of existing mathematical practice and values. In this he has more than delivered on what for me were just at the time suspicions, and I look forward to more productive work in mathematics. An independent exploration of uses of robustness in category theory (developed in a somewhat different way) has recently been published by Corfield (2010).

(3) Also empirically based, but quite different in character from physically or biologically based detectors are those that make intensive use of intentional responses, such as questionnaires and various kinds of more passive survey data. Derived from these are the kind of multi-dimensional index construction utilized in the social sciences, and still nowhere better expounded than in Campbell and Fiske's classic 1959 paper: "Convergent and Discriminant Validation via the Multi-trait Multi-method Matrix". This mode has not been discussed here, but it also appears to be more treacherous ground.[4] (But see Trout 1998, who urges and uses robust methodologies in the human intentional sciences).

Each of the first two of the above categories have seen expansive and creative discussion substantially increasing both the reach and qualifications on the use of robustness concepts both in this volume and in the intervening years in the literature, mostly initiated by rediscoveries of Levins' work by Odenbaugh (e.g., 2006) and Weisberg (2006)[5].

In that survey I did not consider some of the social structures and interactions that give robustness, from multiple eyewitness testimony to the structure of juries although I have explored some of them in teaching. Reproducibility in different labs (which are never exactly the same conditions (Waters 2008)) and with different preparations secures some robustness, as (ideally) does peer review—though the

[4] Twenty years after Campbell and Fiske (1959), that paper was the most cited in the journal, which asked Fiske to review the 2000+ citations it had received to assess the role the methodology might have had in generating progress. Most citations were obligatory "field" reviews that cited but made no use of it, and the success from the much smaller number that did discuss or use it was very disappointing. (The original was in the domain of personality theory, which has been a minefield for attempts at "objective" or "valid" classification scales, so perhaps it would have been better in other areas.) I wondered whether a particularly severe problem in the studies reviewed was in separating the phenomenon being investigated from the tools of analysis, and Fiske agreed (in conversation, 1982). We usually have clearer conceptions where one leaves off and the other begins in the biological and natural sciences, but when this fails, it should be problematic for any robustness analyses.

[5] In Wimsatt (1991), I discuss how multiple views (in this case, 4 different graphical representations of a chaotic phenomenon) were required to understand it. These views are not independent (they are analytically related), but the different representations are crucial to being able to see different aspects of the phenomenon. This is a kind of visual robustness, made relevant by the limitations of our visual and cognitive apparatus.

latter may more often serve to indicate things that should be done to check or to investigate robustness than confirm it. Similar things are represented in the decision to assemble investigators with diverse experience in a single lab to work on a class of problems, and different styles of investigation in different disciplines or different national styles in different countries may lead to investigation of a phenomenon under different conditions that we need to understand the larger picture. Thus it is probably no accident that the character of the *lac*-operon which gave us the first insight as to the mechanisms of gene control and gene expression, would have been disentangled in France, where there was much more emphasis on the physiological dimensions of genetics than in the structurally dominated traditions in England and biochemical hegemony in the U.S. (Burian and Gayon 1991; Morange 1998).

I also considered "pseudo-robustness"—cases where the presumed independence of means necessary for robustness was compromised in ways that were not obvious to the investigators, and where later investigations could show that the arguments for robustness were unsound. The most striking case of this I discussed at length in 1980b—the fallacious arguments against group selection, where the ineffectiveness of group selection appeared to be a robust result across 12 mathematical models. These were shown by Wade (1978) to depend upon 5 false simplifying assumptions. I argued then that the culprit was a common set of heuristics in reductionistic model building, and proposed a systematic corrective in the form of inter-level cross checking of assumptions. (Thus things that looked plausible at a lower level that made simplifying assumptions about the environment could be more easily seen to be unrealistic when one went up one level and considered what were the major relevant organizational features of the environment. And of course, for symmetry, one had better check at lower levels to see that the properties supposed in the model at a given level are indeed robust under the conditions supposed). It seemed to me then, and does so even more now, that demonstrating independence, or characterizing its range and limitations (which also indicates the range of robustness claims), and analyzing any systematic biases in the perspective of existing models, however caused, is a crucial and difficult activity, and indicates again that robustness is no silver bullet. Also left out of my characterization at the time was any attempt to characterize the strength of robustness claims in ways that indicates that different links or paths might have different force or bearing on the central claim said to be robust. This has since been pursued by others—e.g., Stegenga (2009), and by Soler in her introduction to this volume.

3.2 Material Robustness

The methods discussed so far can all be characterized as "inferential robustness." They do not primarily touch on the robustness of phenomena or behavior of objects, except peripherally. In that original sweep of the family of methods relating to robustness, I reached as far as multi-level mapping between states (such as one might find between micro-states and macro-states in statistical mechanics, or molecular and classical characterizations of the gene),[6] and the processes involved in looking at matches and mismatches between descriptions for things that needed to be changed to improve the fit and to develop interlevel explanatory accounts. Rich connectivity has played a crucial role in the development of inter-level mechanistic explanations in genetics and elsewhere (Culp 1994, 1995; Wimsatt 1992).

Robustness and reliability have been deep and entrenched biological design principles in nature throughout evolutionary history. It is not surprising that we should naturally use multi-modal calibration, checking, and inference as an adaptation. Not only does it improve the reliability of our inference and action in the world, but the world itself would not be one in which complex organization could have evolved but for a rich fabric of robust and stable interactions on different size and time scales. This was a central focus of my analysis of levels of organization in 1976, 1994, and 2007b. The analysis of levels of organization using robustness is particularly interesting for philosophy because it would appear that with it, one is getting interesting ontological claims out of broader empirically grounded relationships. Philosophers have of late tended to deny bases of metaphysical claims in empirical relations, no matter how general.

In attempting to analyze what a level of organization was, I moved away from the then universal tendency to discuss such topics as reduction as a relation between theories, which came in levels, and argued that we needed to understand reductive explanations in terms of relationships between phenomena, objects, and regularities at different levels of organization (themselves characterized in terms of robustness). Theories came in levels because levels in nature were the loci of multiple regularities and stabilities among objects and relationships, and one could get a "bigger bang for a buck" by building theories about the entities and relationships one found there (Wimsatt 1976, 1994, 2007a). This would lead naturally to an investigation of material robustness of properties and phenomena in systems, rather than to robustness in our inferential means of determining their properties. This (and the natural robustness of objects in our world) could have been seen as primarily to indicate why it was adaptive that we use robust inference. The discussion in the 1981a paper focused primarily on robust inference rather than robustness in material objects, although my earlier discussion in 1980a considered an important case study, the robustness of chaotic dynamics in ecological systems.

In the intervening years, attention has turned much more intensively to investigations of robustness of natural objects, or of behaviors or properties of natural objects. This has turned out to be particularly useful, as Tom Nickles' (Chapter 14) work has indicated, in the analysis of networks, which might to a prior generation have been seen as rather problematic objects. (Thus Campbell went to great pains to argue that social groups could be seen as entities in terms of the robustness of their

[6] The relation between molecular and classical conceptions of the genes is heterogeneous on both sides, and much more complex than between micro-state and macro-state descriptions in classical mechanics—roughly because there are many mechanisms between micro- and macro- in genetics and many more kinds of anomalies possible (see, e.g., Beurton et al. 2000).

boundaries—generated in terms of their properties as a social network.) With the development of theory, by Watts and Strogatz, Barbasi, and others, for describing the connectivity properties of networks and their consequences, it has emerged in these analyses that some properties of these networks are both interesting and robust. This work in turn can be seen as a conceptual development in a lineage pioneered earlier by Rosenblatt (1958) in ensembles of idealized neural networks (his "perceptrons") and the later development of connectionist networks in the mid-1980s. Both looked for properties common throughout the ensemble of networks meeting certain conditions, and the work beginning in 1969 by Stuart Kauffman looking or self-organizing or generic properties of gene-control networks (things like mean length of cycles as a function of number of nodes and connection density) has this same feature.

3.3 A Central Biological Example: How Is Sex Possible?

A major driver of this interest has been to understand the architectural features characterizing biological organisms that allowed them to be so tolerant of and able to maintain behaviors or states or their characteristic features across generations in the face of environmental or genetic perturbations. I describe here what I take to be the most interesting (and indeed most focal) problem involving robustness in biological organisms. I have taught this case for nearly 30 years as a general puzzle for the architecture of genotype-phenotype mapping that needed solution (Wimsatt 1987, 2007b, Chapter 10), and was long surprised that it was not discussed. This has changed—it is now recognized as important, and I regard it as at least partially solved by the discussions of Wagner (2005). This is what I would call "the paradox of sex", or

"How is sex possible?"

The developing organism is systematically tuned (as a matter of design) so that small differences can have effects on a variety of size scales including the very large, in which context dependence of effects is a common phenomenon, but where it is crucial that most differences do not have significant effects most of the time. Thus the organism can be very responsive to small genetic differences or differences in environmental stimuli or resources, but it is crucial that these not disturb crucial remarkably stable and regulatory species specific developmental and behavioral patterns. (Those used to inter-level relations of the sort characteristic of classical statistical mechanics, where "law of large number" averaging is a reasonable mode of moving from one level to the next, will find the complex interplay of sensitivities and regularizing equilibrations of the relations between genotype and phenotype to be quite remarkable.)

This is a kind of "sloppy gappy" robustness, which can be full of exceptions and where what is required is just that the desired result occur sufficiently frequently—the normal state of affairs in evolutionary contexts. This is because selection works on performance that is only sufficiently regular. Because of the complexity and context dependence of the desired states, crisp regularities are unattainable, but also unnecessary.

Consider the following:

(1) We are given the genetic variability at many loci (of the order of 5%) characteristic of virtually all species of organisms. With a genome size of 25,000 genes, this would mean that 1250 genes are segregating. With just 2 alleles per locus, (since $2^{10} \sim 10^3$) this yields on the order of 10^{375} possible genotypes, and this is an underestimate, since there are often more than 2 alleles per segregating locus.
(2) We also have the scrambling effects of genetic recombination, so that each offspring is essentially without precedent in the specification of its genotype. Offspring of the same parents (save for identical twins) should characteristically differ at many hundreds of loci or more.
(3) Furthermore, we know that small genetic changes can and often do have large effects, and that interaction between genes in producing their effects is the rule rather than the exception. Indeed, characterized biochemically, almost all interactions are epistatic or non-linear in their effects.

    Given these three facts, it is remarkable that any regularity in heredity ensues at all. It would be plausible to expect almost no correlation in phenotypic properties between different members of a species (within the range of properties defined by that species), or between parents and their offspring, and especially to expect frequent lethal interactions. But this would render evolution impossible.
(4) Yet offspring commonly inherit their parents' traits, as well as their fitnesses—not perfectly, but much better than random. The frequency of spontaneous abortions among women of peak reproductive age (25–30 years old) has widely divergent estimates but appears to be somewhere between 15 and 60% (with the higher estimate arising from sub-clinical abortion events undetected by the mother).[7] Many of these involve chromosomal anomalies, with predictably severe results for missing or added chromosomes. But just given the normal genetic variability, even without chromosomal abnormalities, how come mortality is not 99% or more for a normal genomic complement of alleles that have never appeared together? How come all of the nonlinear interactions don't just produce gobbledegook?

For evolution to be possible, there must be heritability of fitness, and to adapt successfully in different environments, it must be possible to select for diverse sets of characters giving adaptation to those diverse environments. This requires not only the heritability of fitness but the heritability of characters or character arrays. Both require the general stability of the species-specific phenotype at many

[7] There are surprisingly few papers estimating this, and fewer recently. Roughly 15% of eggs fail to implant. Estimates of the frequency of first trimester abortions varies substantially between 15 and 30%. Boue et al. (1975) is the most common citation.

levels. Not only must elephants breed elephants, humans humans, and *Drosophila Drosophila*, but the variability and systematic and independent inheritance of individual survival-relevant characters from parents to offspring within each species must be preserved—not glued together with a thicket of epistatic and linkage interactions, or commonly shattered when they are broken up—if temporally and spatially local adaptation to changing environments is going to be possible. We are constantly told of cases where a single base change in a gene or a single amino acid change in a protein has enormous consequences for adaptation and function at a variety of higher levels of organization. But this must be the exception rather than the rule for evolution as we know it to be possible. (Sickle-cell anemia remains the classic case here, and there still aren′t many cases known as yet, though these should increase with our knowledge of developmental genetics.) Nonetheless, the plain fact remains that most genetic changes that happen under biologically normal conditions have no readily discernible effects. (See Lewontin 1978 on "quasi-independence"—the ability to select for one character in evolution without transforming or dragging along a number of other characters, and Wimsatt 1981b for further discussion.)

Wagner (2005) surveys robustness in biological organization in multiple adaptive systems at multiple levels, from the mappings in the genetic code (both the redundancies, and the non-random higher frequency of 3rd position mutations to another amino acid that preserves hydrophilic or hydrophobic interactions at that position) up through various cellular and developmental processes showing multiple pathways and canalizations. I believe that these multiple piecemeal robustnesses and canalizations go a long way to explaining a deep puzzle: the possibility of the heritability of characters and fitness in sexual species with normal degrees of genetic variability. Since there is significant genetic variability in natural populations, a great deal of epistatic (non-linear) interaction between them, and sexual recombination produces genetic combinations that have never arisen before, the heritability of characteristics is a mystery. But given the surveys of robustness discussed by Wagner, in effect, we are exaptively prepared for sexual recombination with the causes of robustness arrayed in a diffuse and distributed manner.

But these may just as well be an effect as a cause, or over time, co-causal with resultant population genetic dynamics. Livnat et al. (2008) urge that under some conditions, in sexual species, the ability of an allele to perform well across diverse genetic backgrounds will be selected for, and model this alternative arrangement. They conclude that selection for this would be especially strong in transient conditions—i.e., those with mixing of populations having a diversity of alleles some of which do not mix well. So it would seem that starting with enough exaptive robustness for sex to be viable (as noted by Wagner), this process should drive evolution to a state in which there is widespread heritability of fitness of alleles across backgrounds.[8] Livnat et al. also note that this process would increase additive

[8] This raises questions about Wagner's assumption that environmental fluctuations should be a stronger driver than genetic compatibilities. As I noted in (2006) he considered only mutation and ignored recombination, which could produce variances that would be orders of magnitude higher than those produced by mutation—though of course not necessarily larger than environmental variance.

components of fitness in such alleles, a possibility I raised in 1981b, p. 146, as a product of selection for the arguably equivalent “quasi-independence” of characters noted as a requirement for evolution by Lewontin (1978).

There are interesting possible parallels here with the evolution of standardization for machine parts, or cross-platform compatibility of software in the evolution of technology, or training for professionals that require certain competencies. In all three cases, things that transport better to other contexts should be selected for, and this applies not only to things like absolute compatibility but to convergent pressures to adopt a common interface, where the cost here is in the steepness of the learning curve for divergent applications.

To return to the biological case, we have robustness apparently increasing evolvability. Wagner addresses the apparent conflict here: shouldn’t robustness decrease pressures and potential for evolution by increasing stability of the phenotype and not showing alternative states to selection? His answer is to argue that if robustness, by allowing the accumulation of greater redundant complexity, also generates a larger state space, with more variability that may be expressed under different conditions, then robustness increases the possibility of normally unexpressed variability that may emerge under new conditions. This plausibly involves an increase in evolvability. In so doing he is supposing that robustness—despite first appearances—is context sensitive, urging something that Nickles has emphasized in his contribution for this volume.

3.4 Qualifications on Robustness

Tom Nickles’ suggestion that robustness should actually be regarded as a 4 place predicate: “system S is robust to perturbation p in degree d except where c” seems exactly right. These or some such similar qualifications seem appropriate. Any piece of detection machinery can obviously break down in ways that are suggested by its architecture, and any tool must be used in a way appropriate to its organization. And any specific kind of linkage will be sensitive to some kinds of stimuli and perturbations and insensitive to others, and usually, to various degrees. Material robustness calls for a study of these qualifications. I argued above for a kind of statistical robustness in adaptive organization which is “sloppy gappy” driven by the characteristics of selection processes. But there are others.

There are some potential ambiguities in the arguments Tom has accepted (from Doyle et al. 2005) that selection for increases in robustness will tend to increase complexity, and that in turn will increase fragility. He has referred to a common occurrence in the evolution of technology. But here we need to distinguish between potential fragility and actual or probable failure. While it is true that adding new linkages or mechanisms to a system to increase robustness can also introduce new failure modes, this does not mean that probability of failure would be increased over the range of conditions where such a system is normally used. The characterization I gave of robustness was purely topological—it pointed to the ways in which alternative parallel paths would increase reliability if there were no changes in the reliability of component linkages, either due to their architectural redesign or to base reliabilities in the components. Changes in these may increase or decrease overall reliability.

Anyone driving a car 50 years ago could expect far more failures than they would now, even though with the increased complexity of a car's operating control systems and their dependence on micro-processors, we have given up even the ability to identify necessary repairs, much less make them, although to someone having the appropriate computer diagnostic tools that utilize all of the imbedded sensors in the cars' control systems, the chances of correct diagnosis may be significantly increased. Obviously there are now many more ways in which such a system can fail, but the net reliability and usability of a computer-controlled injection and ignition system (improving cold weather starting and eliminating hot weather vaporlock) is apparent to anyone who has to start a car on a cold morning. Note that here we have increased the robust operating range of the engine—something often found with continual design refinements. Finer machine tolerances have decreased oil consumption, while computer controlled ignition and valve timing, and the introduction of 4 valves per cylinder have increased combustion efficiency and reduced fuel consumption while increasing specific power output. Synthetic oils have increased the time between oil changes and the mean lifetime of engines has increased. While it is true that the use of electronic components means that they can fail in several new ways (they could be burned out by the electromagnetic discharges induced by a nuclear explosion,[9] there are new possibilities for programming errors (apparently the cause of anomalous acceleration found in 2008 in a variety of Toyota automobiles) or hardware wiring errors. Indeed, the increasingly smaller size of integrated circuits generates an increasing chance of cosmic-ray induced errors. For all of this, the net effect has been significant increases in the reliability, tractability, and efficiency of running engines.

And this reticulate complexity also infects inferential robustness in complex experimental and detection systems. All of this is suggestive of the reticulate complexity of nested analytical and control procedures documented by Lena Soler (Chapter 10) in her rich and beautifully detailed discussion of analytical and experimental robustness in the demonstration of weak neutral currents. But the very complexity allows inferences that would not have otherwise been possible. So is this increased fallibility, or robustness over a greater performance range? Probably both.

What has happened with the technology of automotive engines is that the engines in question have evolved so that they would now be characterized by a state-space with more dimensions. If at a molar level, they still have the same number of functional systems however, we can say that there is a mapping from the later larger

[9] When I worked at the National Cash Register Company (NCR) in 1962 designing adding machine components, there was a sudden interest in hydraulic logic computers, which mimicked electron flows with water streams, exhibited by the defense department. These would be isomorphic to electronic ones (with diodes, adders, and registers), but would not be sensitive to electromagnetic fields induced by nuclear weapons. Their multiple other disadvantages ruled them out however.

number of variables of the newer state space into the smaller number of variables characterizing the functional organization in the older one. And the net lumped failure probabilities in the older (or lower resolution) state space may have decreased even though there are now more ways to fail. So while increases in complexity may generate new kinds of failure modes in systems, this is quite consistent with net increases in overall system robustness. In addition, of course, if we learn more about the system and how its links interact with each other and with inputs to the system, we may better understand its proper calibration and limitations of its use. How much of this also affects the instrumentation and experimental checks of the apparatus for detecting weak neutral currents?

3.5 Robustness and Entrenchment

Finally, robustness, canalization, or other forms of stabilization in biological organization, whether genetic or environmental, and however secured, should provide a primary target for generative entrenchment. Generative entrenchment of an element is a state in which the action or presence of that element has many and diverse consequences for the further development of the phenotype or system in which it is embedded. Things that are deeply entrenched are things that are very conservative in evolution because their change or disruption has a high probability of having far reaching negative fitness consequences. In an evolutionary process, changes are made that act in ways that are modified by the existing dynamical structures in the system. For heritability of effect, mutations would be favored that plug into elements of the developing system that are stable—i.e., robust elements or behaviors, and other things build on these. As a result increasing chains of dependencies are constructed that act in ontogeny, and the more deeply entrenched elements acquire a longer history. Also as a result entrenchment (as "pleiotropy") becomes a powerful tool to analyze developmental architectures, and is widely used to construct phylogenies. So robustness acquires indirectly yet another critical role in the analysis of biological systems. But this is another story (Wimsatt 2001, 2007b, Chapter 7).

Acknowledgement I wish to thank Lena Soler for her imaginative conception of a conference on robustness, for inviting me, and for the productivity of her resulting vision and Tom Nickles for continuing productive interactions over many years.

References

Beurton, P., R. Falk, and H. Rheinberger, eds. 2000. *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge Studies in Philosophy and Biology. Cambridge: Cambridge University Press.

Boue, J., A. Boue, and P. Lazar. 1975. "The Epidemiology of Human Spontaneous Abortions with Chromosomal Abnormalities." In *Aging Gametes. International Symposium, Seattle*, edited by R.J. Blandau, 330–48. Basel: Karger.

Burian, R.M., and Jean Gayon. 1991. "Un Evolutionniste Bernardien à l'Institut Pasteur? Morphologie des Ciliés et Evolution Physiologique Dans l'œuvres d'André Lwoff. " In

L'Institut Pasteur: Contribution à Son Histoire, edited by M. Morange, 165–86. Paris: Editions la Découverte.

Calcott, B. 2010. "Wimsatt and the Robustness Family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings." *Biology and Philosophy*, preprints (2-10-2010).

Campbell, D.T. 1966. "Pattern-Matching as an Essential in Distal Knowing." In *The Psychology of Egon Brunswik*, edited by K.R. Hammond. New York: Holt, Rinehart, and Winston.

Campbell, D.T., and D.W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56:81–105.

Corfield, D. 2010. "Understanding the Infinite I: Niceness, Robustness, and Realism." *Philosophia Mathematica* (III) 18:253–75.

Culp, S. 1994. "Defending Robustness: The Bacterial Mesosome as a Test Case." In *PSA 1994*, vol. 1, edited by D. Hull, M. Forbes, and R.M. Burian, 46–57. East Lansing, MI: Philosophy of Science Association.

Culp, S. 1995. "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles." *Philosophy of Science* 62:438–58.

Downes, S. 1992. "The Importance of Models in Theorizing: A Deflationary Semantic View." In *PSA-1992*, vol. 1, edited by D. Hull, M. Forbes, and M. Okhrulik, 142–53. East Lansing, MI: The Philosophy of Science Association.

Doyle, John, et al. 2005. "Robustness and the Internet: Theoretical Foundations." In *Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies* (Santa Fe Institute Studies in the Sciences of Complexity), edited by Erica Jen, 273–85. Oxford: Oxford University Press.

Feynman, R.P. 1967. *The Character of Physical Law*. Ithaca, NY: Cornell University Press.

Glymour, C. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Grant, P. 1999. *The Ecology and Evolution of Darwin's Finches*, rev. ed. Princeton, NJ: Princeton University press.

Hudson, Robert G. 1999. "Mesosomes: A Study in the Nature of Experimental Reasoning." *Philosophy of Science* 66:289–309.

Kaufmann, S.A. 1969. "Metabolic Stability and Epigenesis in Random Constructed Genetic Networks." *Journal for Theoretical Biology* 22:437–67.

Levins, R. 1966. "The Strategy of Model-Building in Population Biology." *American Scientist* 54:421–31.

Levins, R. 1968. *Evolution in Changing Environments*. Princeton, NJ: Princeton University Press.

Levins, R. 1974. "The Qualitative Analysis of Partially Specified Systems." *Annals of the New York Academy of Sciences* 231:123–38.

Lewontin, R. 1963. "Models, Mathematics, and Metaphors." *Synthese* 15(1):222–44.

Lewontin, R. 1978. "Adaptation." *Scientific American* 239(3):157–69.

Livnat, A., C. Papadimitriou, J. Dushoff, and M. Feldman. 2008. "A Mixability Theory for the Role of Sex in Evolution." *PNAS* 105(50):19803–8.

Morange, Michel. 1998. *A History of Molecular Biology*. Cambridge, MA: Harvard University Press (English translation of *Histoire de Biologie Muleculaire*, 1994. Paris: Editions la Découverte.)

Odenbaugh, Jay. 2006. "The Strategy of 'the Strategy of Model Building in Population Biology'." *Biology and Philosophy* 21:607–21.

Raerinne, J. 2010. "Generalizations and Models in Ecology: Lawlikeness, Invariance, Stability, and Robustness." PhD diss., University of Helsinki, Finland.

Rasmussen, N. 1993. "Facts, Artifacts, and Mesosomes: Practicing Epistemology with the Electron Microscope." *Studies in History and Philosophy of Science Part A* 24(2):227–65.

Rosenblatt, F. 1958. *Principles of Neurodynamics*. Washington, DC: Spartan Books.

Stegenga, J. 2009. "Robustness, Discordance, and Relevance." *Philosophy of Science* 76(5):650–6621.

Trout, J.D. 1998. *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*. New York: Oxford University Press.

Van Valen, L. 1989. "Three Paradigms of Evolution." *Evolutionary Theory* 9:1–17 (July).

Wade, M.J. 1978. "A Critical Review of the Models of Group Selection." *Quarterly Review of Biology* 53(2):101–14.

Wagner, A. 2005. *Robustness and Evolvability in Living Systems*. Princeton, NJ: Princeton University Press.

Waters, K. 2008. "How Practical Know-How Contextualizes Theoretical Knowledge: Exporting Causal Knowledge from Laboratory to Nature." *Philosophy of Science* 75:707–19.

Weisberg, M. 2006. "Robustness Analysis." *Philosophy of Science* 73:730–42.

Weisberg, M., and K. Reisman. 2008. "The Robust Volterra Principle." *Philosophy of Science* 75:106–31.

Willinger, Walter, and John Doyle. 2005. "Robustness and the Internet: Design and Evolution." In *Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies* (Santa Fe Institute Studies in the Sciences of Complexity), edited by Erica Jen, 231–71. Oxford: Oxford University Press.

Wilson, E.O., and W. Bossert. 1971. *A Primer of Population Biology*. Stamford, CT: A. Sinauer.

Wimsatt, W.C. 1976. "Reductionism, Levels of Organization and the Mind-Body Problem." In *Consciousness and the Brain*, edited by G. Globus, I. Savodnik, and G. Maxwell, 199–267. New York: Plenum.

Wimsatt, W.C. 1980a. "Randomness and Perceived-Randomness in Evolutionary Biology." *Synthese* 43:287–329.

Wimsatt, W.C. 1980b. "Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy." In *Scientific Discovery-Vol. II: Case Studies*, edited by T. Nickles, 213–59. Dordrecht: Reidel.

Wimsatt, W.C. 1981a. "Robustness. Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M. Brewer and B. Collins, 124–63. San Francisco, CA: Jossey-Bass Publishers.

Wimsatt, W.C. 1981b. "Units of Selection and the Structure of the Multi-Level Genome." In *PSA-1980*, vol. 2, edited by P.D. Asquith and R.N. Giere, 122–83. Lansing, MI: The Philosophy of Science Association.

Wimsatt, W.C. 1987. "False Models as Means to Truer Theories." In *Neutral Models in Biology*, edited by M. Nitecki and A. Hoffman, 23–55. London: Oxford University Press.

Wimsatt, W.C. 1991. "Taming the Dimensions–Visualizations in Science." In *PSA-1990*, vol. 2, edited by M. Forbes, L. Wessels, and A. Fine, 111–35. East Lansing, MI: The Philosophy of Science Association.

Wimsatt, W.C. 1992. "Golden Generalities and Co-opted Anomalies: Haldane vs. Muller and the Drosophila Group on the Theory and Practice of Linkage Mapping." In *The Founders of Evolutionary Genetics*, edited by S. Sarkar, 107–66. Dordrecht: Martinus-Nijhoff.

Wimsatt, W.C. 1994. "The Ontology of Complex Systems: Levels, Perspectives and Causal Thickets." *Canadian Journal of Philosophy* supplementary volume #20, edited by Robert Ware and Mohan Matthen, 207–74 (reprinted as Chapter 10 of my 2007b.)

Wimsatt, W.C. 2001. "Generative Entrenchment and the Developmental Systems Approach to Evolutionary Processes." In *Cycles of Contingency: Developmental Systems and Evolution*, edited by S. Oyama, R. Gray, and P. Griffiths, 219–37. Cambridge: MIT Press.

Wimsatt, W.C. 2006. "Reductionism and its Heuristics: Making Methodological Reductionism Honest." *Synthese* 151:445–75.

Wimsatt, W.C. 2007a. "On Building Reliable Pictures with Unreliable Data: An Evolutionary and Developmental Coda for the New Systems Biology." In *Systems Biology: Philosophical Foundations*, edited by F.C. Boogerd, F.J. Bruggeman, J.-H.S. Hofmeyer, and H.V. Westerhoff, 103–20. Amsterdam: Reed-Elsevier.

Wimsatt, W.C. 2007b. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge: Harvard University Press.

Chapter 4
Achieving Robustness to Confirm Controversial Hypotheses: A Case Study in Cell Biology

Emiliano Trizio

4.1 Introduction

Wimsatt's robustness scheme[1] and Nickles' notion of multiple derivability[2] should be seen as attempts to capture the logical structure of a variety of strategies aimed at strengthening the reliability of a scientific result. Within experimental sciences, multiple derivations are to be preferred to a single one, and producing multiple independent derivations, whenever possible, is widely regarded as the most effective way to warrant the conclusions of a scientific enquiry. In these disciplines robustness is sought for mainly as a means to counter the epistemically potentially harmful consequences of the fact that something counts as evidence in favor or against a hypothesis only under a number of assumptions, whether explicit or implicit. Some of these assumptions constitute the general theoretical background of the experiment, some others, less general or even contextual, are necessary for justifying the correctness of the experimental technique and for the interpretation of its result. Hence, there is always room for the doubt that the experiment has produced pseudo-evidence deriving from wrong assumptions and that what is attributed to the object of the enquiry is but an artifact created by the experimental activity. Wimsatt's scheme, as it stands, describes the logical structure of the fortunate and wished-for situation in which a multiplicity of different procedures based on independent theoretical assumptions yield the very same result. Arguably, it is unlikely that two or more utterly different techniques should be essentially flawed and yet agree *by chance* on one result out

[1] Wimsatt (1981), Reprinted in this volume, chapter 2.

[2] Nickles (1989).

E. Trizio (✉)
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France

Archives Husserl, Paris, France

Department of Philosophy, Seattle University, Seattle, WA, USA
e-mail: emilianotrizio@hotmail.com

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_4,

of countless possible ones.[3] Robustness can thus become, to some extent, a remedy for the highly indirect and theory-laden character of scientific evidence. However, as it is to be expected, one thing is an idealized methodological scheme and its in-principle logical grounding,[4] but quite another is its actual implementation in specific examples of real scientific practices. That is why cases studies can cast light on several sides of the issue. In what follows, I will focus on two aspects of the role of robustness in contemporary laboratory science by analyzing a case of ongoing research on bacterial endocytosis in mammalian cells, in which robustness, in the sense of multiple convergent derivations of the same result, is explicitly targeted. A group working at the *Institut Pasteur* in Paris is presently carrying out this research. The first aspect to be investigated is that robustness has to be situated in the dynamic context of the debates surrounding a specific research. The fact that in principle multiple independent derivations enhance the degree of confirmation of a hypothesis does not tell us much about how the scientific community, in the context of a given debate, values the quest for robustness. As we shall see, robustness plays a specific role whenever the available evidence is insufficient to settle a scientific controversy. More precisely, robustness, in this case, will appear to constitute an essential and explicit requirement, for the result to be established contradicts standard scientific views. The second aspect that I will investigate is the relation between the very structure of Wimsatt's logical scheme and actual scientific practices. In particular, I will focus on the theoretical independence between derivations and on the identity of the results. It will appear that real situations offer, in general, a more complicated pattern in which a multiplicity of derivations are indeed combined, but their independence comes in degree and the results they yield stand with one another in a relation of partial overlap rather than identity. The analysis will proceed as follows: in Section 4.2 I will recall some basic notions about endocytosis; in Section 4.3 I will reconstruct how recent findings have brought into question some standard views about endocytosis, and in what way robustness has been seen as a strategy to settle the issue. This will help us putting robustness in the dynamic context of an actual scientific research. Section 4.4 will present an analysis of the interplay between multiple means of determination in this particular case. Different schemes will be introduced to account for partial independence of the experimental techniques and the complex relations existing among their results. Moreover, these examples will be discussed in the framework of Nederbragt's taxonomy of experimental strategies.[5]

[3] A classical argument of this type is to be found in Hacking (1985). Although this article will focus on the notion of theoretical independence only, it is worth mentioning that a careful historical analysis may in principle reveal that a convergence between two or more theoretical techniques has been at least partly due to a mutual adjustment between them perhaps taking place at the level of their implementation or of the interpretation of their results. Whenever this is the case, the techniques in question would still be *theoretically independent* while not being *genetically independent* (see Chapter 1, Section 1.8), and this would render problematic the chancy character of their convergence.

[4] For the different interpretations of which see Chapter 9.

[5] Nederbragt (2003).

In the *conclusion* I will sum up the results obtained throughout the analysis and suggest that the simple logical situation portrayed by Wimsatt's robustness scheme is often to be regarded as an aim to be pursued through a long and stepwise process or even as a regulative ideal directing the researchers' efforts, rather than as a readily available option in their methodological tool-box.

4.2 Theoretical Background of Endocytosis

The plasma membrane, i.e. the interface between a cell and its environment, can both uptake nutrients and communicate with other cells of its environment in virtue of a variety of mechanisms. Ions and small molecules (e.g. amino-acids and sugar) can enter the cell through a membrane channel or via the action of membrane pumps; macromolecules and larger objects (even viruses and bacteria) are internalized via a process termed endocytosis. The study of endocytosis is of crucial scientific significance for (1) investigating the relation between a single-cell organism and its environment, (2) casting light on the evolution of cells, (3) understanding the way in which a complex organism, somehow conceived as a large colony of cells, constitutes a unity in virtue of frequent interaction among its cells (for instance, endocytosis occurs in the signaling among neurons), (4) understanding the viral and bacterial invasion of cells and the resulting pathologies.

There are several different types of endocytosis. Figure 4.1 presents a recent taxonomy of these mechanisms in mammalian cells.

This taxonomy bears a great significance for the present case study, for it is based on the size of the cargo being internalized by the cell: indeed each specific type of endocytosis appears to act for cargos whose size varies within a certain range. Endocytosis processes fall into two main categories: phagocytosis (cell eating) and pinocytosis (cell drinking). Both phagocytosis (the uptake of large particles conducted only by specific types of cells) and macropinocytosis (occurring with cargos whose size does not exceeds 1 μm) involve the formation of membrane extensions via the assembly of actin filaments, see Fig. 4.1. The remaining three types

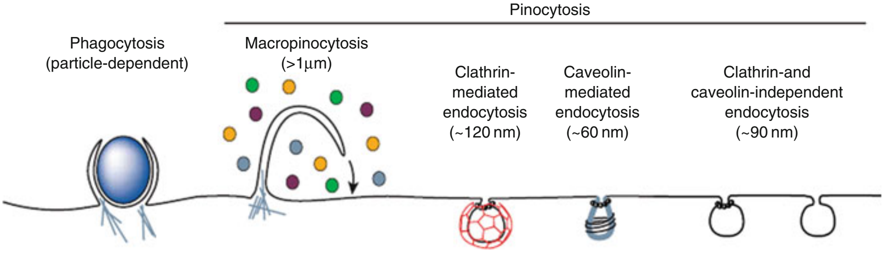

Fig. 4.1 A taxonomy of endocytic mechanism in mammals. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Conner, S.D., and S.L. Schmid "Regulated Portals of Entry into Cell", 422:37–44 © 2003. http://www.nature.com/nature/journal/v422/n6927/abs/nature01451.html

of endocytosis take place, according to this taxonomy, only at smaller scales, and require the formation of various types of vesicles making possible the uptake of external material.

Let us focus on clathrin-mediated endocytosis (CME), which is the object of the researches analyzed in this case study. As it is indicated in Fig. 4.1, clathrin-mediated endocytosis was believed to occur only when the size of the cargos does not exceed 120 nm (or perhaps 150 nm), and at a far smaller scale than those internalized via phagocytosis and macropinocytosis. Thus, it was believed that clathrin did not play an important role when bigger external objects enter the cells. Figure 4.2 schematizes as an example the subsequent steps of CME.

Clathrin coated vesicles are formed via the invagination of the cell membrane. The vesicles act as vehicles carrying cargos of external material into the cell. There are three main actors in the formation of clathrin-coated vesicles: (1) clathrin triskelia, (2) adaptor protein complexes AP, and (3) a GTPase called dynamin. Clathrin is originally present in three-legged structures called triskelia. In vitro studies have showed that clathrin triskelia can assemble into polygonal cages without the intervention of other factors.[6] However, within the cell environment, that is, under physiological conditions, clathrin triskelia assemblage into polygonal coats requires the presence of another constituent: the heterotetrameric adaptor protein complex

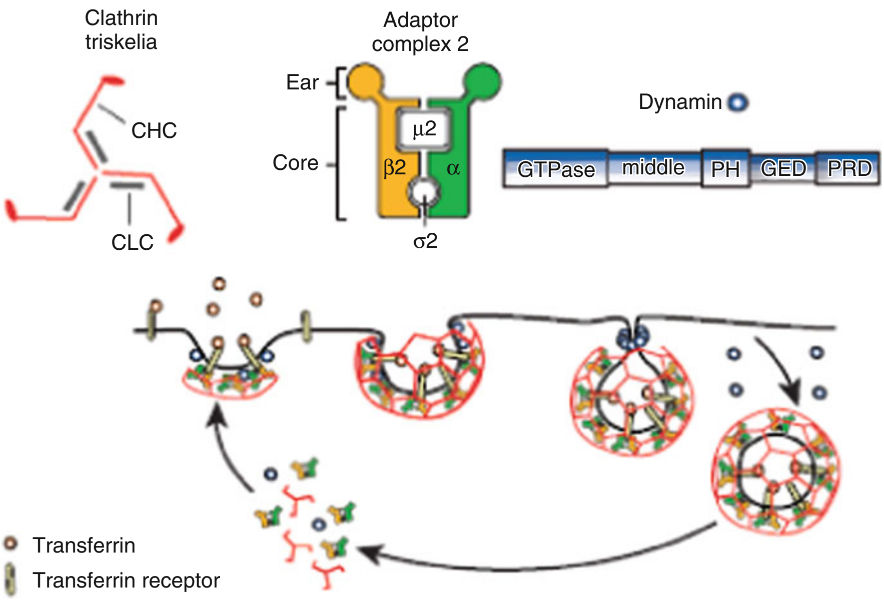

Fig. 4.2 A scheme of clathrin-mediated endocytosis (CME). Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Conner, S.D., and S.L. Schmid "Regulated Portals of Entry into Cell", 422:37–44 © 2003. http://www.nature.com/nature/journal/v422/n6927/abs/nature01451.html

[6] Keen (1987, p. 1997).

called AP2. As Conner and Schmid say, clathrin, due to its capacity to create curved lattices, acts as the "brawn" determining the invagination of the membrane, while the protein AP2 acts as the "brain" directing the assemblage of clathrin.[7] The third main actor in CME, the GTPase dynamin, plays a central role in the endocytic vesicle formation by forming collars around the neck of a coated invagination and controlling its fission from the membrane (perhaps acting as a sort of pinchase severing the neck[8]). Finally, the coat of the vesicle disassembles and the cell can reuse its constituents.

4.3 Some Recent Findings Concerning Clathrin-Mediated Endocytosis and the Conflict with the Dominant Views

Having reconstructed the dominant views about endocytic processes, we are now in the position to better appreciate the disruptive effect of some recent experimental results. We have already seen to what extent the study of endocytosis can enhance our understanding of the mechanisms of bacterial and viral infection, for both bacteria and viruses exploit the machinery of entry into cells in order to invade them and reproduce. In an article published in 2005[9] Esteban Veiga and Pascale Cossart, two researchers working at the *Institut Pasteur* in Paris, announced that invasive bacteria (for instance, *Listeria monocytogenes*) enter cells in clathrin-dependent manner. This study was conducted with the technique of fluorescence microscopy, which allowed filming the different phases of the invasion of mammalian cells by *Listeria*. The striking character of this research becomes apparent if only one thinks that *Listeria* is a bacillus reaching 2–6 μm in length, that is up to twenty times more than the supposed upper limit for CME. It has also been claimed that ligand-coated tags whose size varies between 1 and 6 μm can be internalized in a clathrin-dependent way. Consequently, if these results were confirmed, it would appear that the taxonomy presented in Fig. 4.1 is fundamentally wrong, for clathrin-mediated endocytosis would take place even at larger scale than macropinocytosis! Further theoretically independent evidence for Veiga and Cossart's conclusions derives from indirect biochemical tests:

> The possible role in bacterial entry of the major proteins involved in the clathrin-dependent endocytosis (...) was tested by small interfering RNA (siRNA), which resulted in great reduction in bacterial entry. Moreover, the relevance of other components of the endocytic machinery (...) was shown by siRNA knock-down. Decreased expression of these proteins strongly inhibited *L. monocytogenes* entry, demonstrating a major role of the clathrin-dependent endocytic machinery during *L. monocytogenes* infections. (Veiga and Cossart 2006, p. 502)

---

[7] Conner and Schmid (2003, p. 40).

[8] But see Kelly (1997) for doubts about this hypothesis.

[9] Veiga and Cossart (2005).

In other words, it was possible to induce a measurable decrease in bacterial entry by reducing, via biochemical means, the ability of the cell to produce clathrin. This is, of course, an indirect test that is based on theoretical principles different from those underlying any kind of microscopy.

It is also clear that, if *Listeria* can invade mammalian cells in a clathrin-dependent way, clathrin must be able to assemble to form far larger vesicles than previously imagined. Indeed, Veiga and Cossart consider the architecture of the large clathrin structures observed around entering bacteria as "the key issue that remains unsolved".[10]

The fact that the old taxonomy of endocytic process was deeply entrenched and almost taken for granted is highlighted by the skepticism with which the scientific community has reacted to these findings. The article describing the role of clathrin in the endocytosis of *Listeria* was not accepted by the journal *Nature*, whose referees demanded further independent evidence and, in particular, evidence based on electron microscopy. In other words, the results obtained by means of fluorescence microscopy alone were not deemed to confer enough reliability to the result, even though, as we have seen, biochemical tests provided them with a theoretically independent, albeit indirect, support. Crucially then, the demand of multiple derivability was an explicit one. The reason for this skepticism is two-fold: on the one hand, as we shall soon see, fluorescence microscopy does not provide enough details about the role and architecture of clathrin in this kind of phenomena, and, on the other hand, further independent derivation was judged necessary to enhance the reliability of the result. In sum, multiple derivability appears as an explicit criterion for the acceptance of scientific results, precisely because the results in question are in sharp contrast with the dominant views about these phenomena.

Before turning to the analysis of the experimental strategies adopted to enhance the reliability of the experimental results, it is worth trying to understand what exactly the nature of this scientific controversy is. What does support the claim that clathrin-coated vesicles cannot exceed 120–150 nm?

An analysis of part of the relevant literature[11] indicates that there is no "no go theorem" showing the impossibility of large clathrin structures put forward by biochemists. Moreover, the authors of the publications do not explicitly aim at showing that such structures do not exist. The situation is rather the following: in vitro cryo-EM and X-ray studies performed since the 1980s have never detected such large structures. This has motivated a general agreement on the validity of the previously mentioned taxonomy of endocytosis processes and, consequently, a number of research projects that do not take into account the possible role of clathrin in the internalization of large objects. Coherently enough, also the possible co-involvement of clathrin and actin has not been taken into account, precisely because

[10] Ibid., p. 503.

[11] See Zaremba and Keen (1983), Keen (1987), McMahon (1999), Marsh and McMahon (1999), Conner and Schmid (2003), Ehrkich et al. (2004), Veiga and Cossart (2005, 2006), Veiga et al. (2007), Cheng et al. (2007).

actin was considered to intervene only in the internalization of large particles, which requires the formation of membrane extensions. Figure 4.1 clearly illustrates this belief, for the presence of actin filaments is indicated only in the case of phagocytosis and macropinocytosis, while it now seems to intervene also in CME. The absence of an explicit and articulated theory about the size of clathrin-coated vesicles probably explains why Veiga and Cossart talk of a "dogma" that has to be challenged and not about an established theory. This "dogma" has become deeply entrenched among cell biologists, and the available evidence about the role of clathrin in the internalization of large objects is not strong enough to justify a revision of the currently accepted taxonomy.

In the next paragraph I will try to characterize the logical structure of the research being carried out in order support Veiga's and Cossart's claim.

4.4 The Experimental Strategies Implemented to Achieve *Robustness*: A Type of *Robustness Scheme* and Its Peculiar Features

The research presently being conducted at the *Institut Pasteur* in Paris is based on the joint implementation of different types of microscopic techniques. In this specific case, the researchers study *Listeria* infection by combining fluorescence microscopy with electron tomography. There are four different techniques (two variants of fluorescence microscopy and two variants of electron microscopy) that can be combined in a variety of ways in order to enhance the reliability of the results. Of course, each of them might introduce different artifacts due to the specific sample preparation. The two types of fluorescence microscopy are:

FM_1 The sample is prepared with chemical fixation + biochemical techniques. The plasma membrane is then permeabilized; primary antibodies detect clathrin molecules and secondary antibodies, carrying the fluorescent tags, adhere to the primary antibodies. The cells are dead and only static images can be obtained. In this case, it is the procedure of the chemical fixation and permeabilization that is suspected to produce artifacts.

FM_2 The cells are kept alive, and made to express clathrin plus fluorescent tags by means of a modification of the cell's DNA. As the cells remain alive throughout the observation, this technique allows making videos representing the infection dynamic sequence. The potential artifacts affecting this technique might be due to the fact that the experimenters have intervened on the very mechanism controlling the expression of fluorescent clathrin. The clathrin thus obtained may not behave exactly in the same way as the ordinary one. In technical terms, the artifacts may result from the *non-naturality* of clathrin.

These two variants of fluorescence microscopy have already been implemented to find evidence for the claim that *Listeria* can enter cells in a clathrin-mediated way. The researchers' efforts now focus on the implementation of two types of electron microscopy:

EM_1 Classic electron microscopy (room temperature): chemical fixation, dehydration and resin embedding. Possible labeling with antibodies carrying gold particles that are attached to other antibodies, which can in turn adhere to the sample. Once more, the cells are dead, and only static images can be obtained. Potential artifacts are due to the fixation technique and to the labeling.

EM_2 Cryo-fixation followed by cryo-electron microscopy: in order to maintain the cells in their native state, the sample is frozen alive. The cryo-fixation is performed in such a way that the formation of crystals is avoided, for crystals might damage the sample. At the moment, it is not yet clear what kind of artifacts might affect the results of this advanced version of electron microscopy, which has the advantage of avoiding the potential interference of chemicals used for fixing the sample and allows the observation of the real mass distribution.

One might think that these four techniques in addition to biochemical tests would be used to achieve an experimental configuration of the classical robustness type. A look at how the situation would appear, if this were the case, will help us to explore the specificities of this example.

There are several reasons why the experimental procedures applied to the study of clathrin cannot be represented in the way Fig. 4.3 does, and I will now examine some of them in detail. Indeed such a robustness scheme is often inadequate to represent the real complexity even of a simple experimental investigation, let alone of one involving sophisticated combinations of techniques. To begin with, the arrows indicating the different attempted derivations of the result do not signal in any way

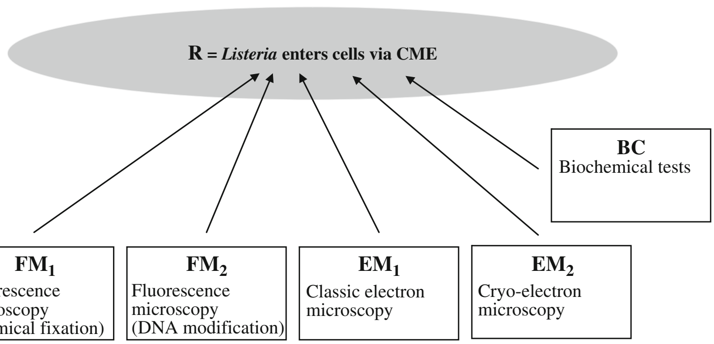

Fig. 4.3 An inaccurate picture of the experimental researches on CME

the reliability of each of them taken per se, and hence the amount of evidential support it can confer to the result, regardless of the more or less strong convergence with its fellow derivations. This is not, however, the problem I will try to investigate here.[12] I will instead focus on two related issues: the independence of the different derivations on the one hand, and the actual relationships among their supposedly identical results on the other. In order to discuss the first problem, I will resort to the classification of types of experimental reasoning proposed by Hubertus Nederbragt[13] in a study mainly devoted to a methodological analysis of research about bacterial invasion. Nederbragt suggests a hierarchy of three different types of reasoning: (1) reliable process reasoning, (2) variation of independent methods, (3) multiple derivability.[14] To each level of the hierarchy there corresponds a step forward in the confirmation of a local hypothesis. The first level consists in a thorough and systematic check of all the steps included in an experimental procedure whose reliability or application is being questioned. Systematic errors and possible artifacts are taken into account, but only minor modifications of the experimental protocols are performed at this stage. In the case of the bacterial invasion experiments, one might, for instance, modify "the cultivation times of mammary gland cells before the bacteria are added".[15] This stage of theory confirmation is ubiquitous in science and aims at enhancing the solidity of a single derivation, and, as I said, will not concern us here. The second strategy consists in performing small modifications of the experimental procedure without changing the theoretical principles on which it is based. In the case of microscopy, an example of this type of variation is given by the adoption of a different fixation technique for the sample. To appreciate the difference between (1) and (2), one should recall that no experimental procedure is entirely fixed or completely specified, and that even without changing any piece of equipment, or technique of data analysis, the experimenters can vary certain parameters of the experimental protocol (just as in the previously mentioned example of the cultivation times). The second level of the hierarchy, instead, implies the deliberate attempt to modify the protocol, for instance, by using different materials or procedures, but without resorting to techniques that are based on different theoretical principles. Only at the third level do we find the implementation of two or more techniques that are based on different theoretical bodies of knowledge. The difference between (2) and (3) can be better understood by considering that while every modification of an experimental procedure implies a certain change in the overall set of background theoretical hypotheses underlying it (in principle, also a modification taking place at the level of reliable process reasoning), only at the third level of the hierarchy do the experimenters turn to experimental techniques

[12] For an analysis of the intrinsic solidity of a single derivation and its impact on the overall resulting robustness, see Chapter 10.

[13] Nederbragt (2003).

[14] In the following chapter of this volume, Nederbragt will further develop his taxonomy by adding a fourth level (triangulation), which I won't discuss here (See Chapter 5, Section 5.2.3).

[15] Ibid., p. 609.

whose *theoretical principles* are entirely different. In keeping with our example, one thing is to change a fixation technique while implementing the same type of microscope, and quite another is to switch, say, from fluorescence microscopy to electron microscopy. In the latter case, the very theoretical principles of the experimental techniques are completely different.

The reliability of a hypothesis increases with the subsequent application of each method.[16] It is noteworthy that Nederbragt does not equate multiple derivability with robustness, for, after presenting this hierarchy of strategies, he claims: "At each level, robustness is the result of a dynamical process of reasoning and theory making"[17] and, already in the discussion of the example of the mesosome, we find the following statement: "Both, multiple derivability and variation of independent methods produce robustness but on different levels."[18] Robustness is therefore intended as a generic property of reliability/solidity obtainable with different strategies among which by far the most effective in enhancing the reliability of a result is multiple derivability. Robustness could be seen, by arguing along these lines, as a synonym of "invariance under *different* derivations". Indeed, even the minor changes adopted at the level of reliable process reasoning can show the invariance of a result under these changes.[19] Now, in this analysis, while adopting Nederbragt's hierarchy of experimental reasoning, I will talk of "robustness" only when a number of convergent derivations based on different theoretical principles are performed, that is in the situation corresponding to Nederbragt's "multiple derivation" (level 3). This choice is more coherent with Wimsatt's terminology and will help us to highlight the specificity of the present case study. Further, given that, at the second level of the hierarchy, one cannot really talk of theoretically independent derivations in the sense I have specified above, I will call the second level "variation of experimental techniques" instead of "variation of independent methods".

By applying Nederbragt's classification to our case study, we immediately recognize a major inadequacy of Fig. 4.3. That scheme treats on an equal footing, as convergent derivations, two variants of fluorescence microscopy, two variants of electron microscopy, and the biochemical tests. This might not be incorrect in principle, but in this way we are left without an indication of what derivations are based on different theoretical principles and jointly build up robustness proper. It is clear that only groups of derivations of the type fluorescence microscopy/electron microscopy/biochemical tests could potentially yield a real case of robustness in the sense previously specified, for between the two variants of fluorescence microscopy on the one hand, and between the two variants of electron microscopy on the other,

[16] According to Nederbragt, the hypothesis of the mesosome (a supposed cytoplasmatic organelle of bacteria, whose existence seemed to be proved by observations carried out with electron microscopy) was abandoned because it failed to pass tests at level 2 and 3.

[17] Nederbragt (2003, p. 609).

[18] Ibid., p. 606.

[19] The idealized situation of an experiment being repeated several times exactly in the same way could be seen as a limiting unrealistic case showing independence with respect to space, time and perhaps to the experimenters.

there is only a difference in the preparation of the samples. In short, Fig. 4.3 does not allow us to tell apart what groups of derivations can be considered as belonging to level 2 of Nederbragt's classification (variation of experimental techniques in my terminology) or to level 3 (multiple independent derivations). I will therefore adopt the convention of indicating the groups of derivations constituting only variations of experimental techniques with parallel arrows. Bundles of arrows with different directions will correspond in turn to groups of derivations based on different theoretical principles. In short, robustness will correspond only to the convergence of arrows having different directions, which is tantamount to saying that, from the point of view of a robustness analysis, the parallel arrows indicating the variants of a fundamental type of derivation are seen as belonging to an equivalence class and can be thus, in a sense, identified.

Complexifying the classic robustness scheme in the light of Nederbragt's classification helps us dealing with the second problem too, namely that of the identity of the results obtained with the various derivations. Let us first notice that different techniques often yield completely different experimental outputs. For instance, a colored film realized with fluorescence microscopy will have, at first sight, little in common with a black and white picture obtained by means of an electron microscope. In most cases, the identity is achieved only at the level of the judgments expressing the final results based on the interpretation of the experimental outputs.[20] This is one of the main reasons why the convergence of different experimental techniques can be (and often is) questioned by different groups of experimenters. This case at hand, however, shows that the experimenters cannot limit themselves to comparing the interpretations of the various experimental outputs in the hope of bringing to light their convergence (or even identity), because the different experimental techniques do not all render accessible the same aspects of the object of the enquiry. Let us first consider fluorescence microscopy. Both variants of this technique have the great advantage of intervening on clathrin itself by rendering it clearly detectable. In one case antibodies carrying fluorescent tags adhere to clathrin molecules, in the other case the cell is made to express a modified form of clathrin carrying fluorescent tags. It follows that the presence of clathrin is clearly signaled by a certain color in the images thus obtained. However, owing to the limited resolution power of the currently available fluorescence techniques, these images have a very limited resolution. At the moment, they can only reveal the presence of clathrin where and when bacteria invasion takes place (localization of clathrin), but say little about the role of clathrin and the shape it assumes. Hence, there is no way, on the basis of these images alone, to conclude that the bacterium is actually recruiting clathrin to invade the cell, and that the entry mechanism involves a certain architecture of clathrin triskelia. Unlikely though it might seem, the presence of clathrin may be due to other phenomena occurring during the bacterial invasion. On the other hand, purely biochemical methods do provide evidence for the causal role of clathrin

[20] I will not consider here cases in which images can be seen themselves as final results of an experiment. For an analysis of this interesting situation, see Chapter 7.

in this type of endocytosis, but fail to give information about its localization and shape. Biochemical tests are in a sense "blind", and can only allow the observation of more or less strong functional correlations between the knock-down of certain RNA sequences and the number of cells that are infected by *Listeria*. The best technique available at the moment is doubtlessly electron microscopy, for it can produce extremely high-resolution images. If successfully implemented, this technique would give at once information about the localization and the shape of the clathrin structure. The observation of a clathrin structure enveloping an entering bacterium and the realization of a series of images corresponding to each major stage of the internalization would in turn constitute evidence for the causal role of clathrin.

This being the situation, we realize that we are not in the presence of a number of different derivations of the result *R*, as one might have thought at first sight. Of course, one should also avoid the opposite mistake of thinking that the different techniques simply provides disjoint bits of knowledge about the phenomena in question, which the experimenters would only need to put together as pieces of a puzzle. If this were the case, no robustness at all would be achieved. I will argue that the real logical structure of these experimental researches can be grasped only by considering the successive partial overlaps of different results. For this purpose, besides the *total result R* (= Listeria invade mammalian cells via CME), we have to consider two partial results (1) the *core result* that is the mere localization of clathrin where the invasion takes place, (2) the *intermediate result* that is the causal role of clathrin in the invasion. For the purposes of this analysis, we can assume that the second partial result implies the first, for it is implausible that clathrin should play a causal role in the entry of *Listeria*, without being localized around it. In short,

$$R \rightarrow R_{\text{int}} \rightarrow R_{\text{core}}$$

Where:

R_{core} = clathrin is localized where *Listeria* enters the cells
R_{int} = clathrin has a causal role in the entry of *Listeria*
R = clathrin exerts this causal role by assuming a certain shape (= Listeria enters cells via CME)

Having reformulated the result in this way, we see that FM_1 and FM_2 establish R_{core}, BC establishes R_{int} and hence also R_{core}, and, finally, EM_1 and $EM_{2,}$ were the current researches to be successful, should be able to achieve the total result.

There is, though, a further complication making the case at hand more intricate (and more inadequate the scheme presented in Fig. 4.3). *In practice*, it is extremely hard to use EM_1 and EM_2 without the aid of either FM_1 or FM_2, because clathrin is too small a molecule to directly identify and only when it assembles in large basket structures, can we identify it by the shape and size of the basket.[21] However, in this

[21] Clathrin coated vesicles have been imaged with EM for many years.

specific case, we still ignore what the size and shape will look like. As we have seen, an advantage of FM techniques consists in the fact that they modify clathrin itself, either by adherence of antibodies (FM_1) or by intervention on the DNA sequences controlling the production of clathrin (FM_2). Thus, the colorful images and videos obtained by means of FM are, at present, the best available means of *localizing* clathrin structures. In contrast, an experimenter using EM alone would have to be so lucky as to size an image of a bacterium at the very moment in which it enters a cell: countless attempts would thus be necessary. Furthermore, the architecture of clathrin surrounding it would have to show a characteristic and recognizable aspect. It is for these reasons that experimenters resort to the technique of correlative microscopy, which involves the joint implementation of FM and EM. For instance, a sample is first prepared with chemical fixation for FM_1. The fixation technique implemented is a sophisticated one, suitable also for electron microscopy. The resulting images allow recognition of clathrin assembling and, only at that moment, the sample is prepared for observation with EM_1 (case $FM_1 + EM_1$) or frozen for observation with cryo-electron microscopy (case $FM_1 + EM_2$). Of course, in this way, the potential artifacts of technique FM_1 may influence the results of electron microscopy too. That is why the researches crosscheck the results thus obtained with the two other possible correlative techniques: case $FM_2 + EM_1$ and $FM_2 + EM_2$. It is clear that, in this way, any potential artifact due to one of the four initial techniques would not affect the results of at least two of the correlative techniques thus obtained.

Summing up the results of this analysis, it becomes now possible to propose a scheme representing the experimental configuration that the researchers of the *Institut Pasteur* are trying to obtain.

In Fig. 4.4, the thick arrows stand for relations of logical entailment between results, whereas the thin arrows indicate, as usual, the different experimental derivations. I have adopted the previously introduced convention according to which parallel arrows represent the derivations based on the same theoretical principles, whereas major differences in theoretical principles are indicated by changes in the direction of the arrows. It is important to stress that the four correlative techniques that should eventually converge on *R* do not by themselves build up any robustness proper, because their results are all based on the body of knowledge supporting EM. This is the reason why they are represented by parallel lines. In other words, correlative microscopy does not provide multiple independent derivations.

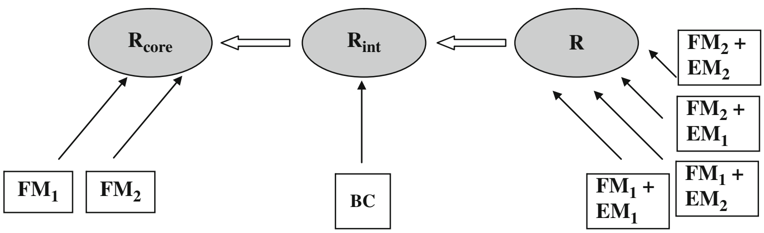

Fig. 4.4 A scheme of the experimental configuration that researchers are currently trying to obtain

A certain number of considerations follow from the analysis of this scheme. In the first place we can appreciate to what extent the joint implementation of different experimental techniques in order to strengthen the reliability of a result can differ from the idealized scheme of multiple independent convergent derivations. In this case, as neither FM nor BC can establish the total result R, the experimenters would not obtain the total result R in a number of independent ways *even if* they succeeded in realizing the situation portrayed by Fig. 4.4. Nevertheless, one would be wrong in thinking that robustness does not play a role here, for the derivations of a result are *ipso facto* derivations of the sub-results entailed by it. From Fig. 4.4 we can therefore extract, by shifting all the arrows towards the left, a classic robustness scheme involving three independent types of derivations of the partial result R_{core} (see Fig. 4.5).

To be sure, the convergence of FM and BC on R_{core} is already achieved, while the convergence of correlative techniques remains, at present, a hope. If this research succeeded, also R_{int} (that is the causal role of clathrin) would be a robust result obtained with two independent techniques (see Fig. 4.6).

By decomposing the total result in different sub-results, we come to understand that, in the course of an investigation, the experimenters work, whether consciously or not, on different superimposed layers of result enjoying different degrees of reliability. It may not be easy, as in this case, to develop a full multi-modal access to the phenomenon under scrutiny, but there can still be a more or less extended core of robustness lying at the heart of the final comprehensive result.

One might object that the techniques based on correlative microscopy rely also on the workings of FM, and that, therefore, Fig. 4.5 is incorrect in that it assumes that FM derivations (on the left) and correlative derivations (on the right) are mutually independent. This objection, however, can be countered by pointing out that the role

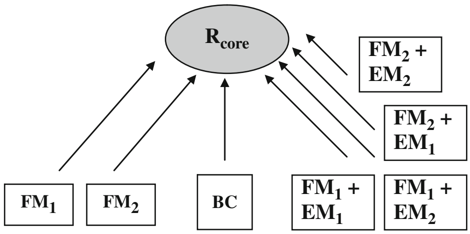

Fig. 4.5 A classic robustness scheme involving three independent types of derivations of the partial result R_{core}

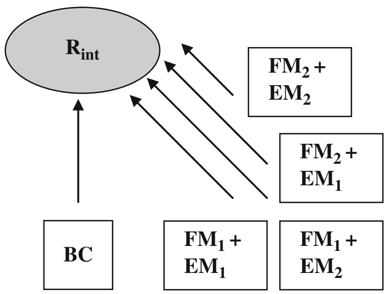

Fig. 4.6 A classic robustness scheme involving two independent types of derivations of the partial result R_{int}

of FM in the correlative techniques is a merely auxiliary one. FM is used only to localize clathrin, and in no way interferes with the images obtained by means of an electron microscope. The sample preparation, as we have seen, is performed in two completely different manners, thus rendering extremely unlikely that an artifact due to both preparation techniques may affect the result. Finally, as I said, only practical considerations lead the experimenters to avail themselves of FM as a "pointer" for electron microscopes, given that there is no reason in principle preventing them for trying to spot an entering bacterium by means of EM alone. At present, there are indeed other teams of researchers trying to locate clathrin baskets with a huge number of random observations based on EM, and their researches might lead to an experimental configuration structurally equivalent to that portrayed in Fig. 4.5, but without the auxiliary role of FM.

4.5 Conclusions: Robustness As a Methodological Attractor

The present study has focused on a specific case of scientific research in which experimenters explicitly aim at the achievement of robust results, that is of results that can be derived in different theoretically independent ways. Once this research is situated in the dynamic context of the debates surrounding it, it appears that the strategy of looking for multiple independent derivations becomes an explicit demand when a new result challenges the views previously enjoying the consensus of a scientific community. This is due to the fact that the amount of reliable evidence needed for the establishment of a scientific result (in the sense of "establishment" implying the consensus of the scientific community) is proportional to the extent to which the result in question calls for a revision of pre-existing bodies of knowledge. Presumably, the reaction of the scientific community to the new findings about CME would have been far warmer, if these findings had not been in contradiction with widespread and almost taken-for-granted beliefs.

With the analysis of the underlying logical structure of this ongoing and still open-ended research in cell biology, I have tried to show to what extent the pattern governing the quest for robustness can be more complicated than the simple situation described by a multiple convergence of independent derivations of the same result. This complexity is likely to be widespread across different scientific disciplines, and the neat, idealized convergence described by the classical robustness scheme may well be an exception rather than the rule, for different techniques often cannot yield the very same result. Nevertheless, the quest for robustness still motivates the researchers' efforts. In this specific case, the decomposition of the final aim of the research into partial results has shown that robustness can be achieved at the level of a core-result, more easily derivable in a variety of independent ways. But this is not yet the end the story. Fluorescence microscopy is still unable to produce detailed images of the entry of bacteria, but, at present, several attempts are being conducted to improve the technique so as to gain a significantly higher resolution.[22]

[22] See, for instance, Betzig et al. (2006).

Should these technical advances be successful, it might be possible, one day, to obtain more detailed optical images of bacterial invasion, which could be usefully compared with those obtained by means of electron microscopy. Results that are richer and richer in content could be thus derivable in different independent ways. In this sense, robustness can be characterized as a “methodological attractor”, an idealized logical scheme guiding the efforts of experimenters and technicians, and at the same time, playing a crucial role in the acceptance of a result by the scientific community.

Acknowledgements Writing this chapter would have been impossible without the keen help of Dr. Anna Sartori Rupp of the *Institut Pasteur* (Paris), who has patiently introduced me to the current state of the research on endocytosis and whose suggestions have constantly guided my work. I wish also to thank Léna Soler and Hubertus Nederbragt for helping me to improve this article.

References

Betzig, E., et al. 2006. “Imaging Intracellular Fluorescent Proteins at Nanometer Resolution.” *Science* 313:1642–5.

Cheng, Y., et al. 2007. “Cryo-Electron Tomography of Clathrin-Coated Vesicles: Structural Implications for Coat Assembly.” *Journal of Molecular Biology* 365:892–9.

Conner, S.D., and S.L. Schmid. 2003. “Regulated Portals of Entry into Cell.” *Nature* 422:37–44.

Ehrkich, M., et al. 2004. “Endocytosis by Random Initiation and Stabilization of Clathrin-Coated Pits.” *Cell* 118:591–605.

Hacking, I. 1985. “Do We See Through a Microscope?” In *Images of Science*, edited by P.M. Churchland and C.A. Hooker, 132–52. Chicago and London: The university of Chicago Press.

Keen, J.H. 1987. “Clathrin Assembly Proteins: Affinity Purification and a Model for Coat Assembly.” *The Journal of Cell Biology* 105:1989–98.

Kelly, B. 1997. “Is Dynamin Really a ‘Pinchase’?” *Trends in Cell Biology* 7:257–9.

Marsh, M., and H.T. McMahon. 1999. “The Structural Era of Endocytosis.” *Science* 285:215–20.

McMahon, H.T. 1999. “Endocytosis: An Assembly Protein for Clathrin Cages.” *Current Biology* 9:R332–5.

Nederbragt, H. 2003. “Strategies to Improve the Reliability of a Theory: The Experiment of Bacterial Invasion into Cultured Epithelial Cells.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 34:593–614.

Nickles, T. 1989. “Justification and Experiment.” In *The Uses of Experiment. Studies in the Natural Sciences*, edited by D. Gooding, T. Pinch, and S. Schaffer, 299–333. Cambridge: Cambridge University Press.

Veiga, E., and P. Cossart. 2005. “Listeria Hijacks the Clathrin-Dependent Endocytic Machinery to Invade Mammalian Cells.” *Nature Cell Biology* 7(9):894–900.

Veiga, E., and P. Cossart. 2006. “The Role of Clathrin-Dependent Endocytosis in Bacterial Internalization.” *Trends in Cell Biology* 16(10):499–504.

Veiga, E., et al. 2007. “Invasive and Adherent Bacterial Pathogens Co-Opt Host Clathrin for Infection.” *Cell Host & Microbe* 2:1–12.

Wimsatt, W.C. 1981. “Robustness, Reliability and Overdetermination.” In *Scientific Inquiries and Social Sciences*, edited by M.B. Brewer and B.E. Collins 123–162. San Francisco, CA: Jossey-Bass.

Zaremba, S., and J.H. Keen. 1983. “Assembly Polypeptides from Coated Vesicles Mediate Reassembly of Unique Clathrin Coats.” *The Journal of Cell Biology* 97:1339–47.

Chapter 5
Multiple Derivability and the Reliability and Stabilization of Theories

Hubertus Nederbragt

5.1 Introduction

Ever since philosophers of science started to show interest in experimentation the discussion of the inductive processes by which theories may be inferred from experiments has been addressed as well. The claim that experiments may serve justificatory purposes has been defended by Nickles (1989); in his view experimentally obtained phenomena may lead to reasoning to a theory instead of reasoning from a theory to a prediction. This view has enabled new approaches to the study of reasoning in experimental settings.

One of the fields which may profit from such new approaches is that of biomedical science, the field in which diseases are investigated with the use of laboratory experiments that are performed to obtain more knowledge of pathogenesis, diagnosis and therapy of diseases. But although the study of the history and sociology of biomedical sciences is well under way, the study of the epistemology of biomedical sciences is still in its infancy.

As a contribution to ameliorate this situation I adopted Nickles' notion of multiple derivability (1989, p. 308) to illustrate the use of an inductive strategy for improving a theory in a biomedical experiment (Nederbragt 2003). Multiple derivability is an inductive strategy in the process of theory making. It may be defined as the process in which a theory may be inferred from two or more observations or premises that have been obtained by theoretically different and independent methods. According to Nickles

> [r]eliability is enhanced when a result can be derived in more than one way. . . . A single 'proof' is nice, but multiple derivations are naturally preferred in science because they are robust. . . . For example, if very different theoretical models or research techniques yield the same result, we have robustness. This robustness can be of several kinds. . . . In the form of multiple derivability, discoverability itself can be robust.

H. Nederbragt (✉)
Descartes Centre for the History and Philosophy of the Sciences and the Humanities, Utrecht University, Utrecht, The Netherlands
e-mail: h.nederbragt@planet.nl

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_5,

The concept of robustness had been discussed in detail earlier by Wimsatt (1981) who also, in passing, used the terms multiple determination and multiple derivability for this concept; he placed it in the tradition of such concepts as triangulation and consilience of inductions.

The case I described to illustrate the strategy of multiple derivability was that of the experiment of invasion of bacteria into cultured animal cells (Nederbragt 2003). I will return to this experiment later, but the philosophical question of this experiment was how we are justified that seeing an agar plate with bacterial colonies makes us infer that the bacteria have invaded the cultured cells with which the experiment was started.

Remarkably, biomedical students and scientists do not consider this as a great problem. For them it is obvious that making pictures of the cells with an electron microscope is the next step to make the theory that the bacteria have invaded the cells more reliable. Apparently, strategies of multiple derivations seem unproblematic for scientists in an experimental context, but they have not been analyzed sufficiently in a philosophical context. This may warrant the more detailed analysis presented here.

Before I proceed I have to clarify some of the terms I use to prevent possible confusion. My use of the term "multiple derivability" refers to local theories; this is due to my reading of Nickles (1989) who seemed to place it in the practice of experiment (see quotation above). Therefore, for me multiple derivabiliy is a procedure, a way of improving theories (although the term linguistically suggests an end result of the process). The same goes for the term robustness as used by Wimsatt (1981, p. 126) who stated that robustness may involve, among others, problem-solving procedures (although here again the term refers to an end result). For my use of the term robustness I will consider it as the outcome of a strategy: something is robust when it remains invariant under a multiplicity of independent derivations. Thus, as I will show later, the outcome of multiple derivability is a more robust theory.

A similar clarification is necessary for my use of the terms "reliability" and "stability". Reliable and stable may be two different features of a theory. I consider reliability as epistemological stability but some theories may not become stabilized by epistemological arguments alone (e.g. because they are weak) but also by social, i.e. non-epistemological, factors. Thus my use of the term "stabilization" is broader than stabilization by epistemological strategies alone.

This study is divided into three parts, each containing reflections on aspects of multiple derivability that are only loosely connected to each other. The first part will describe and discuss the family of comparable strategies, including multiple derivability, triangulation, and consilience of inductions. The question I will address is whether those strategies bear different names for good reasons, or whether they are basically the same. I analyse them against the context of scientific practice and therefore it will appear that, although all three have been loosely defined as variants of the same principle, in scientific practice they have served different purposes. In the subsequent parts of this paper I will concentrate on multiple derivability and triangulation.

The second part shows several cases, mainly of biomedical origin, which I will use to investigate in more detail what it means when we talk of independent derivation. The cases will also give more insight in the strategy of multiple derivability and show that a theory, that cannot sufficiently be supported by multiple derivability because of lack of independence of derivations, may become reliable due to triangulation. Because in several cases the problem is raised of independence in relation to reproducing evidence, the question of repeatibility and reproduction will be addressed in a separate section.

The third part will address the problem that theories may lack sufficient reliability because of the absence of both multiple derivability and triangulation. Nevertheless they may become stabilized within the science community because of social interaction between method, theory and scientists of the research group. In this part a biomedical research project where this was the case will be analyzed.

5.2 A Family of Strategies

Three strategies have been proposed that may lead us to robustness. Although they belong to a single class of robustness strategies, I will show that these three strategies may be applied to different conditions of theory formation.

The oldest is Whewell's consilience of inductions and is described as taking place "when an Induction, obtained from one class of facts, coincides with an Induction, obtained from another different class. This Consilience is a test of the truth of the Theory in which it occurs" (Whewell 1847, aphorism XIV, p. 469).

A related strategy is triangulation which is discussed by Star (1986) and may be described for the moment, in her words, as the strategy that refers to the use of different methods, taken independently, to describe the same object.

The third is multiple derivability, which I defined as "the strategy by which a theory is supported by the evidence obtained through two or more independent methods that differ in the background knowledge on which they are based" (Nederbragt 2003). I preferred the name because it represented the dynamic principle of a strategy for making a more reliable theory by converging methods.

Now I want to address the question whether different names are justified because they refer to different principles or whether the three names are only different names for the same principle.

5.2.1 Multiple Derivability

First I shall analyse multiple derivability in more detail. I use a case from biomedical research to illustrate the features of the strategy. Milk gland epithelial cells of a cow were cultured in vitro as a monolayer in the presence of mastitis (inflammation of the udder)-causing bacteria. The principle of the experiment is that after removing and killing bacteria outside the cultured cells the contents of the cells are plated on agar plates and that the resulting bacterial colonies on the agar make it possible to infer that the bacteria have invaded the cultured cells (Nederbragt 2003). Scientists are aware that this is not a sufficient argument; after the paper of this experiment had been published (Döpfer et al. 2000) I was told by a bacteriologist that another strain of bacteria, cultured on intestinal cells, had been shown to hide under the monolayer of cells instead of inside the cells, which could have been inferred from our experiment as well for our bacteria. Philosophers of science are also aware that the finding of bacterial colonies is not a sufficient argument for a theory of invasion but they call it underdetermination. Thus, the experimental results may make us infer one or more alternative theories, or, in other words, the results may be explained by more theories than the one under study. One of the aims of my paper of 2003 was to show that biomedical scientists, working on bacterial invasion into cultured cells since 1955, have used more than one test to demonstrate this type of invasion; in those cases in which they applied an independent method, different in theoretical and technical background, they used the strategy of multiple derivabilty.

I have now the opportunity to bring forward some additional points that were not discussed in my previous paper.

(a) I introduced electron microscopy (EM) as an independent and different method to show that bacteria had invaded the cultured cells. But as was pointed out to me later,[1] in the case that EM is a *standard*, unambiguous method to show invasion, the colony count method is *validated* by EM and its outcome is made dependent on the EM outcome, and therefore derives its justification as giving reliable evidence from the EM method. After such a justification the only reason to use the method of colony counting is, that it is cheap and simple compared to EM, not that it gives a more robust theory. Thus, no multiple derivability is taking place then. However, it may be argued that the invasion theory derived from EM experiments is also underdetermined, e.g. because from the EM pictures it is also possible to infer that the bacterium is only partly surrounded by cellular material and is in fact still outside the cell. Thus, both methods make us infer the same theory, each of them underdetermined, albeit for different reasons. The scientist may be aware that in both cases alternative explanations may be invoked for his findings, different from the one he was testing. But considering both outcomes of the two different experiments together, i.e. using multiple derivability, a theory of bacterial invasion is inferred that is more robust than each of the two separate theories of bacterial invasion taken in isolation.

(b) In so far as multiple derivability is seen working in my example of cell culture and bacteria, the range of the strategy is limited. It gives robustness to a theory that deals with these particular types of cultured cells and bacteria only and therefore I called it "local". Thus, multiple derivability is a strategy for local theories. This is in contrast to triangulation, which I will discuss in the next paragraph, in which bodies of knowledge may be locked into each other (the term is from Nickles 1989) and thereby may give robustness to theory complexes.

[1] I thank Wouter Meijs for discussing this point with me.

(c) Multiple derivability is a strategy that is seen as useful and even necessary by scientists. Commenting on my explication of multiple derivability a colleague remarked that this is the way science should be done. Biomedical bachelor students reacted spontaneously that EM was the method of choice when the invasion studies with bacterial colonies left room for doubt. In addition I received a long comment (A. Glauert, personal e-mail communication, 2004) on my 2003 publication, describing how electron microscopists proceeded with their studies:

> the scientists .... who started to use the electron microscope ... in biology in the late 1940s, followed the general practice of all types of microscopist[s] and have always used 'the strategy of multiple derivability'. However, we call it 'examination by a number of different methods', and the more the better.

This was followed then by a number of examples of multiple derivability with regard to light microscopy and EM. Therefore, multiple derivability is a strategy that is chosen intentionally, to improve arguments for the certainty of a finding, a hypothesis, a description or a picture, i.e. to obtain more robustness.

Thus, to conclude, multiple derivability is a strategy applied in the practice of improving a local theory by support of evidence from two or more independent different methods.

5.2.2 Triangulation

Although triangulation is a word that is frequently used in passing as if it were clear what is meant, more detailed analyses of what it is are scarce. I will use three publications in which the authors tried to explain in more detail what the features of triangulation may be. I shall treat them in chronological order.

Star (1986) described triangulation as a methodological tool in social and biological sciences, for improving validity and reliability. She gave it two meanings: "it can refer to the *same* method employed independently to describe the same object, or it can refer to the use of *different* methods, also taken independently, to describe the same object". The first use compares independent users; since my aim is to compare triangulation and multiple derivability as a strategy of independent approaches against different background theories, I will concentrate on Star's second meaning only. The aim of her analysis was to "empirically examine the conditions and consequences of combining evidence". She used as a case for this examination the field of brain research at the end of the 19th century in which clinical and basic research were the "realms" from which evidence was combined. Her analysis concentrated on the sociological and historical aspects of this case of triangulation. In the end her conclusion about biases in triangulation was that all participating disciplines, such as physiology and pathology, and the tools with which they were equipped, such as taxonomies and theoretical models, had serious limitations but nevertheless converged to a theory of brain function.

Another, more detailed analysis of triangulation is given by Gaudillière (1994); he uses a case study of the role of cellular transfection studies in the finding of viral and cellular oncogenes. He sketches a network of relations between experimenters and experimental groups in which triangulation takes place. In this process methods and objects are mutually defined. Gaudillière follows Star's view of triangulation as an interplay of actors of different social contexts but expands the concept to a process also occurring in single social space in which experimental cultures have to reach agreement on laboratory strategies. The main purpose and consequence of these triangulations is to deal with uncertainties. He made it clear that it may be possible to distinguish several types of triangulation, depending on the types of actors, sources or situations we have to consider.

The last analysis of triangulation I want to refer to is given by Risjord (2001). It concerns a discussion of the contribution of quantitative and qualitative methods in health care research. Risjord discussed whether triangulation of the two types of methods really give confirmation or whether they are only useful for completeness and inspiration. According to him, opponents to the confirmation view hold that both methods are embedded in different paradigms, meaning that natural (quantitative) and social (qualitative) sciences can not be "blended". Risjord analysed different arguments and concluded that "[a] theory may be composed of a number of questions for which different kinds of methods are suitable. Where this is the case, methodological triangulation will be necessary." The most important feature of triangulation in his analysis is when qualitative data (e.g. data from questionnaires and interviews) are supported by quantitative results (physiological data) and vice versa.

Overviewing the three analyses above it becomes clear that triangulation is the process in which different cultures, realms and disciplines converge on each other, leading to new insights. The insights may include more reliable theories with a wider scope, and confirmations and epistemological stabilizations of such theories when they depend on varying, not necessarily related, research activities. It has a broad application, ranging from laboratory methods (Gaudillière) and methodological cultures (Risjord) to research traditions (Star). It may even lead to new disciplines such as that of in vitro protein synthesis in the context of the development of modern molecular biology (Rheinberger 1997). Despite the broadness of scope which seems to be part of the definition of triangulation as given by the authors (by definition even the results obtained by two different experimenters may triangulate), when looking at the way it is applied by the authors it seems to be mainly used for obtaining convergence of theories in larger theory complexes. Multiple derivability makes local theories, triangulation may make broader theories form the more local ones; multiple derivability goes from methods to theories, triangulation goes from theories to theory complexes.

5.2.3 Consilience of Inductions

Most of the authors who deal with triangulation and multiple derivability give credit to Whewell (1847) for his consilience of inductions. This is the term that Whewell introduced to describe the particular feature of established theories that were reached by the jumping together of inductions from different classes of facts. His emphasis is on "unexpected and wonderful" coincidences which is "one of the most decisive characteristics of a true theory".

Despite its age, the term consilience of inductions is not often applied in philosophical studies of scientific practice; this may have to do with the views philosophers of science bring forward when discussing this strategy. A few examples may illustrate this.

In a paper in which he defended inference to the best explanation Thagard (1978) used consilience of inductions as one of the three criteria for determining the best explanation, the other two being simplicity and analogy. Consilience is a "notion" that

> is intended to serve as a measure of how much a theory explains, so that we can use it to tell when one theory explains more of the evidence than another theory. Roughly, a theory is said to be consilient if it explains two classes of facts. Then one theory is *more* consilient than another if it explains more classes of facts than the other does.

Thagard then went on, as Whewell did, to discuss historical examples of consilience, among them Snell's law of refraction, Lavoisier and oxygen, Huygens and light, the mechanics of Newton and Darwin's theories of evolution. The main point of Thagard's discussion was that consilience of inductions serves to explain with hindsight why, in the history of great theories, one theory was accepted rather than an other.

Laudan (1981), in explicating Whewell's notion of consilience of inductions and trying to clarify it in the context of Whewell's ideas about philosophy of science, described consilience of inductions to occur under the following circumstances:

> 1. When an hypothesis is capable of explaining two (or more) *known* classes of facts; 2. When an hypothesis can successfully *predict* "cases of a *kind different* from those that were contemplated in the formation of our hypothesis" [this quotation is from Whewell]; 3. When an hypothesis can successfully predict or explain the occurrence of phenomena which, on the basis of our background knowledge, we would not have expected to occur.

Laudan admitted that there is some overlap between the three and argued that they have in common their contribution to maximizing the confirmation of a hypothesis. However, in my view the most important point is Laudan's emphasis on the historical character of consilience: "the deduction as to whether a given theory has achieved consilience of inductions can only be reached via a careful study of those other theories with which it is competing at a given time. Without a thorough knowledge of historical context . . . it is usually impossible for us to decide whether a given theory achieves a consilience or not." In this emphasis on historical context Laudan seemed to agree with Thagard.

A third analysis of consilience of inductions I want to refer to is that of Fisch (1985). He wanted to find out what Whewell really could have had in mind when he proposed his consilience and in addition he wanted to examine the intuition that gives us more confidence in a consilient theory. He made a distinction between "the non-consilient case in which a theory is successfully conjectured to jointly explain two correlational generalizations, each over a different class of phenomena" and "the consilient case in which the theory is successfully conjectured as an explanation of one such generalization, and then it surprisingly turns out to apply equally successfully to the other." Fisch states that from the point of view of predictive and explanatory power alone, and in my view we may read "robustness" here, both cases are the same; the difference between the two "would have to lie in their case histories, that is, in the two different processes by which the theory arrived at its final state." Fisch then continued with a detailed analysis of genuine consilience and pseudo-consilience, the latter being characterized by one of the generalizations becoming less different from the other by reinterpretation. The power of the real consilience is that the two generalizations retain their different meanings and this maintenance of difference gives considerable support to the theory.

Although the three analyses are rather different in aim and approach they have in common that they try to formalize what consilience of inductions is and what it does: criterion for theory choice (Thagard), contribution to maximizing the confirmation of theories (Laudan) and the features of real consilience (Fisch), all this in a philosophical context. They also agree in the importance of looking back in the history of theories to make consilience traceable. In addition, they seem to emphasize the deductive power of a theory, whereas triangulation and multiple derivability concentrate on the inductive power of evidence.

This short summary of consilience makes it possible to contrast it with triangulation and multiple derivability.

Consilience of inductions is a strategy to find out what makes theories good theories. Philosophers of science who perform such an analysis use theories made in the past; their interest is how theories became accepted. The daily practice of science is of no direct concern in consilience of inductions.

Triangulation is a strategy used to find out how theories came into being; how, to borrow a phrase of Rheinberger (1997), things became epistemic. The context in which it is applied is that of social interactions of scientists and scientific groups in different cultures and traditions in past and present. However, triangulation also refers to scientific practice. It occurs, or is deliberately applied, when researchers of different fields join forces and combine insights to gain better explanations of seemingly unrelated experimental data.

Multiple derivability describes what practicing scientists do who want to make their theories more reliable. I have called these theories "local" because they are made in a setting of a laboratory or small research group and because they are not yet supported enough to submit in the form of a manuscript to the wider science community. These scientists do not look back, they look forward; multiple derivability is their strategy of choice.

Therefore, concentrating on scientific practice may be more helpful to make distinctions between the three strategies. One of the features of triangulation is that groups or disciplines converge to a robust theory. For consilience of inductions groups do not play a clear role; most authors of analyses of consilience are interested in theories, not the persons, groups or laboratories who found them; they seem to consider history, coincidence and surprise as the main features of consilience. In the concept of triangulation groups, networks and cultures are the actors in the process.

But nothing is defined a priori about the number of people involved and nothing forbids that different cultures try to converge deliberately with the aim to make a theory more robust. As an example the context of the bacterial invasion experiment (Nederbragt 2003) may be used: the PhD student who did these experiments was stimulated to start these because of her work on epidemiological investigations and mathematical modeling from which a hypothesis could be deduced that bacteria may survive within the milk gland epithelial cells. Thus, setting up invasion experiments was meant to make a theory of the presence of bacteria in epithelial cells of the udder more robust. Neither the mathematical modelling nor the invasion experiment were demonstrating that bacteria really could survive inside the milk gland cells of cows, but because of these two converging approaches the theory that bacteria could survive gained more robustness. Clearly, in this example two disciplines converged in one person and triangulation may be seen at work here.

The qualitative and quantitative methods described by Risjord (2001) are also placed in a context of triangulation. In a personal communication (2009) he explained that the two groups using the qualitative and quantitative methods respectively may be characterized as focus groups. The investigations described in his paper used the traditions of both groups intentionally. Accordingly, Risjord called his triangulation "methodological". What is important here is that the examples of triangulation have an aspect of prescription, of a deliberate choice. For this reason I propose that triangulation should be given a place in the hierarchy of robustness strategies that I developed previously (Nederbragt 2003). Depending on the level of theory making I distinguished "reliable process reasoning" (in the experimental set-up), "variation of independent methods" (modifications of procedure within a fixed theoretical background) and finally "multiple derivability" (different methods with different theoretical backgrounds), all three aiming at making a local theory more reliable. I now want to add triangulation as an overarching strategy in which local theories are made more encompassing because of convergence of local theories of different disciplines in one theory or theory complex.[2]

I conclude that, when comparing consilience of inductions, triangulation, and multiple derivability it is justified to consider them as different. Consilience is more directed to the analysis of the historical background of theory choice. Multiple derivability and triangulation are concerned with scientific practice, the latter being the strategy for obtaining robustness of interdisciplinary theories or theory complexes and the former for obtaining robustness of local theories in the experimental environment. This difference in dealing with theories is reflected in the analysis of Thagard (1978), in which he used consilience of inductions as a general, overarching concept in which he analyzed several strategies, including a strategy that may be called triangulation and strategies that may be applied at the level of making reliable methods. I consider these as different when analyzing scientific practice. Since I am

[2] In the same paper I used the case of establishing the function of mitochondria by the joined efforts of electron microscopists and biochemists as an example of multiple derivability; based on the newer views I have of multiple derivability and triangulation I now prefer to classify the mitochondria case as an example of triangulation.

partial to science in daily practice I propose that we concentrate on triangulation and multiple derivability when we try to understand how theories are obtained in the practice of (biomedical) research.

5.3 Reflections on Independence

Above I have defined multiple derivability as being "obtained by theoretically independent methods". In his analysis of robustness Wimsatt (1981) also points out that independence of derivations is an important feature of the process in which a robust result has to be obtained. In this respect, it may be useful to consider what independence means for multiple derivability at work in biomedical experiments. I want to discuss five cases that may help us to get a clearer view of independence of methods and I will also discuss the problem of repeatability of experiments because it may teach us more about what kind of independence we need.

5.3.1 The Case of Perception of an Object

I am sitting at a table and I see a pencil lying on it. Can I trust my eyes? Is that what I see really a pencil? The principle of multiple derivability requires that I infer the same "theory", i.e. I see a pencil lying there, by using another, independent method with a different background theory. I may try my tactile senses. Whereas vision is based on colours, dimensions, form etc. that excitate my retinal cells through the lenses of my eyes, touch is based on texture, dimensions and form (moderately hard material with an elongated hexagonal form) excitating impulses in the nerve ends of my fingers. So there is a difference. Intuitively we may find multiple derivability at work when we do a play of guessing blindfolded the identity of an unknown object. By touch alone we may conclude that the object we perceive is a pencil, and after taking off the bandage we see with relief that the conclusion was right. So multiple derivability is what helped us here. However, we may have been hallucinating. In that case, our whole mind may be confused and since vision and touch are connected by the operations of one brain, both may be affected simultaneously and in the same way when the brain is affected. Therefore, vision and touch may be different, but they seem to be insufficiently independent. However, we may consider them independent when we limit our "theory" of perceiving the pencil to the operation of one mind only, whatever its state.

Wimsatt (1981) discussed such a case to argue that we may regard perception as a criterion for robustness when more than one sensory modality is used to perceive an object; such robustness "is ultimately the primary reason why we regard perception of the object as veridical rather than illusory". In dreams or hallucinations robustness may be obtained in a similar way, but Wimsatt indicates its failure because consistency cannot be obtained through time or across observers.

I conclude from this discussion that robustness is limited to the context in which it is found; the context of my dreams is different from the context of when I am awake and in both situations the reality of an object may be robust because of multiplicity of perception. Therefore derivations may be independent within a certain context, but their independence is also restricted to this context. To establish the reality of objects across contexts, another step of multiple derivability is necessary, at a higher hierarchical level.

5.3.2 The Case of the Investigation of a Crime

A crime has been committed in a house and a person has been accused of the crime because of the following evidence: his handkerchief has been found in the room where the crime was committed, fibers of his clothes were found in the same room and a witness had seen him leave the house at the supposed time of the crime.

The local theory of this case is that the suspect is guilty of the crime. Three types of evidence (observation) are available. The two objects may be identified on chemical analysis of the handkerchief, e.g. DNA investigation, and on forensic physics for identifying the origin of the fibers; they are therefore based on theoretically different methods. The third type of evidence is testimony which is a completely different type of observation, and in view of the discussion in the former section we may conclude that triangulation is the strategy here. However, it can easily be seen that the evidence is unsufficient: the objects may have been left at the site of the crime on purpose by a person who wanted to manipulate the evidence in a certain direction. To make things worse: the witness may have been the person who left those objects at the site of the crime. Therefore, it cannot be concluded that the observations are really independent. This case must be considered as an example of lack of background knowledge and it is of course the purpose of the crime investigators to collect more data to establish the background knowledge that makes it possible to decide on the independence of the observations.

This case and the first one teach us that independence of derivations is valid within and defined by a certain context and that it is necessary to understand the background knowledge of this context to decide how independent the derivations are.

5.3.3 The Case of the Invasion Experiment

Above, I used this experiment for explaining what multiple derivability is. I use it again here, because more can be learned from it with regard to the problem of independence of method and background knowledge. As I explained above, the experiment was directed at demonstrating bacteria in cultured bovine mammary gland cells. To this end the bacteria were cultured for a short time on top of those cells; subsequently all non-invaded cells were killed with an antibiotic and after lysing the still intact cells, in which the bacteria had not been accessible to the antibiotic, the resulting living bacteria were made visible by culturing them on agar on which each bacterium gives rise to a colony. The central question was how we can reliably conclude from a colony count that bacteria had invaded the cultured cells. It appeared then that multiple derivability was the strategy of choice. A literature search (Nederbragt 2003) showed that for the study of bacterial invasion many different methods had been applied in the past: colony counts as in my example above; electron microscopy; light microsopy after staining bacteria with dyes; marking the bacteria with a radioactive label or with a fluorescent antibody. Two or three of these methods were used in each of the studies in which experiments with bacterial invasion were performed. This is most explicitly shown in a study of Hale et al. with Shigella cultured with Henle intestinal cells, described in two subsequent publications (Hale and Bonventre 1979; Hale et al. 1979) in which they used three methods (chemical staining with light microscopy, fluorescent labeling and a colony count method such as ours) for the purpose of multiple derivability, although they described the purpose of the use of these different methods as validation of their methods.

It may be easily seen that the methods I mentioned above are independent of each other. The theoretical background of electron microsopy is in physics: ultrathin slices of cultured cells, perpendicular to the plastic support on which the cells were growing, are cut after rigorous fixation with glutaraldehyde, and the slices are then brought into a beam of electrons; the different degrees of electron density correspond with morphological structures in the cells, including bacteria, and are made visible on a photographic film or on a computer screen. The bacterial colony count is based on microbiological methods and theories regarding the growth of bacteria in colonies on agar plates. The other methods in the literature mentioned above have other background theories, such as (radio)chemistry, optics and immunology in varying combinations. In fact, in biomedical investigations in the laboratory, multiple derivability seems to be a standard strategy for the construction of local theories, but this is a claim that needs to be investigated further.[3] In the end of this paper I will present a case in which stabilization of biomedical knowledge is reached by a different strategy.

5.3.4 The Case of Criteria for Causation in Epidemiology

Epidemiologists have made efforts to make causal inferences from data of epidemiological investigations; generally these data have been obtained from population studies and deal with risk factors and their statistical association with specific

[3] It should be added here that two independent methods may give us more robustness, but sloppy methods, although independent, will not be accepted. Let us, as an example, imagine that an effect of an intervention in an experimental system gives a positive response compared to an untreated control system and that this may lead us to a hypothesis of the system, but the result is not statistically significant; the second, independent, derivation to the same hypothesis is also based on a non-significant experimental result. Although the investigator may feel that she has obtained robustness here she will not be able to get it published in a refereed journal.

diseases, for instance between dietary components and certain types of cancer. Although the associations themselves do not tell us about causal relations between risk factor and disease, considerable efforts have been made to enable a causal conclusion from such data. These have been extensively discussed by Evans (1993) and I will use his "unified concept" of criteria for causation (Table 9.5, p. 174) as my starting point, although I have simplified the list of criteria somewhat for practical reasons. The idea behind these criteria is that the claim for a causal connection between the risk factor and the disease is reinforced each time one of the criteria can be shown to be satisfied in addition to a previous one.

The list may contain the following criteria.

1. *prevalence* of the disease (i.e. the percentage of individuals with the disease in a population) should (in a statistically significant way) be higher in the group of individuals exposed to a putative cause than in the unexposed group.
2. *exposure* to the putative cause should be present more commonly in the group with the disease than in the group without the disease.
3. *incidence* of the disease (i.e. the number of individuals with the disease in a defined time period) should be significantly higher in the individuals exposed to the putative cause than in those that were not exposed.
4. *temporally* the disease should follow the putative cause.
5. a *spectrum* of host responses should follow the putative cause, varying from mild to severe.
6. *elimination* or *modification* of the putative cause should decrease the incidence of the disease.
7. *experimental reproduction* of the disease should occur in higher incidence in man (volunteers) or (laboratory) animals when exposed to the putative cause than in individuals not exposed.
8. the relation between putative cause and disease should be *biologically plausible*.

The criteria 7 and 8, the experiment and the biological plausibility, are independent from the others, mainly because they do not apply to epidemiological analyses but to experimental intervention studies (criterion 7) or are derived (or cannot be derived) from background knowledge (criterion 8) in research performed in disciplines other than those of epidemiology. The main problem of epidemiological data and their analysis in isolation is that, since no intervention is involved, no basis for causal arguments is present.

Two questions are important here: has the idea that the causal claim is reinforced each time one of the criteria is fulfilled anything to do with multiple derivability; and are those criteria independent?

To start with the latter question: when we look at the nature of the research practice from which the criteria may be obtained we find that the criteria 1–6 apply to population studies of spontaneous disease cases or spontaneous exposures. Criterion 6 applies to an intervention study, performed in a population in which an exposure is modified, such as iodine addition to table salt to modify iodine deficiency (a "negative exposure"), as a putative cause of goitre, 7 is a criterion for an experiment under controlled circumstances with animals or volunteers, and 8 is a criterion based on studies in the discipline of pathogenesis, i.e. pathology, cell biology, toxicology, biochemistry or several others, connected to the type of risk factor and the particular disease under study.[4] Since all of the studies related to 1–6 are performed by mathematical analyses of epidemiological databases they should be considered as having low independency, even if different databases have been used; when some of them are combined they still may lead to overestimation of the causal claim. This lack of independence is due to their sharing many methodological and theoretical aspects which may not help to get rid of possible errors in the databases and biases in the outcome of their analysis. In this respect criterion 6 seems to have a higher degree of independence of the other five than these have of each other, because it is a manipulation study instead of a correlation study. But even if a new population survey is performed, say an intervention study by analysis of a database for finding an effect of modifying a lifestyle factor in a population, the theoretical and methodological assumptions are not sufficiently different from those of other surveys to warrant that multiple derivability is a strategy here.

However, when we look at criterion 7 the situation changes considerably when we combine it with one or more of the criteria 1–6. This is because a study with volunteers or experimental animals is performed with the purpose of finding an effect of an alleged risk factor. Assuming that such an effect is obtained it is still not certain that the factor under study is the causal factor in the population too, but when it is combined with a epidemiological study the causal claim is certainly made more likely. However, in my view multiple derivability is not involved here, which answers the first question asked above; I think that we deal here with the combined theoretical evidence of two studies from different fields or disciplines and according to my hierarchical scheme proposed in the former part we see here triangulation at work.

I conclude from this theoretical epidemiological case that multiple derivability in the sense as I described before was not applicable here but that the causal claim may be justified by triangulation.

5.3.5 The Case of the Diagnosis of Acute Myocardial Infarction

A diagnosis is a statement of the condition of an individual. A diagnostic test, and this may be physical examination, a blood test or a CT scan, aims at reducing uncertainty about the condition of the individual. This condition may be a disease state but many other conditions may be tested as well by diagnostic tests. The outcome of a test may have a prognostic aspect, such as when a malignant tumour has been diagnosed, pointing to the presence of a life-threatening situation. An analysis of the epistemology of a diagnostic test may also demonstrate why, when using multiple

---

[4] However, after more critical analysis of the criteria, it may be argued that criterion 8 has more to do with background knowledge and that it is difficult to see it as an independent derivation. Therefore, I shall omit it from further discussion.

derivability as a strategy, independence of derivations is necessary: the inference may lead to overestimation of or unjustified trust in the hypothesis or theory when the different methods are not independent. This will be made clear below.

A diagnosis is made by inductive reasoning and may be considered as an inference to the best explanation. It is inferred from certain symptoms. As such, a diagnosis is a hypothesis that may be tested, e.g. by autopsie after the death of the patient, by a therapy, by waiting what will happen to the patient in course of time, or by any other outcome of a decision taken about treatment or non-treatment of the individual.

How do we know a diagnosis is correct? In a clinical situation this depends on the characteristics of the diagnostic test. All diagnostic tests may result in diagnosing a person with a disease she does not have (false-positive) or diagnosing a person free of the disease although she actually has it (false-negative). New tests should have been validated for their accuracy in a comparative setting in which the true positive of all positive scoring individuals or true negative of all negative scoring individuals are determined, the result being called the sensitivity resp. specificity of the test. Such validation studies relate the outcome of the new test to a so-called gold standard, i.e. the reference test that is supposed to determine the disease state unambiguously.

In practice, tests with a specificity and sensitivity of 1.0 do not exist and therefore a diagnosis made on a single individual may be wrong. In case of doubt of the outcome of the diagnostic test, it may be improved by performing a second test. Because of the risk of overestimation it is easy to see that when the second test has to improve the diagnosis it should be different from the first, both in its methodological and theoretical background. Suppose we have performed a blood test; unless we are interested in the reproducibility of the test result (which does not help the patient) it is not useful to repeat the test with the same blood sample and the same machine because the conditions responsible for the result of the first test have not changed. It is better to take a new blood sample and, in case of doubt of the quality of the test conditions, to use another but similar test system, e.g. in another laboratory. When a diagnosis may have severe consequences a greater difference may be required. A new blood sample is slightly more different than the old blood sample used for the second time; a urine sample is far more different than a new blood sample. Physical examination is completely different from chemical tests on blood or urine, but may much more easily lead to false positive or negative results and can therefore not be used as an alternative. In fact, this problem of how to repeat the diagnosis is similar to the replication problem described by Collins (1992) and I will return to it later.

When two tests, methodologically and theoretically independent of each other, point to the same target condition we have more confidence in the correctness of the diagnosis than with one test only. It is therefore a clear example of multiple derivability. This is the strategy which is chosen when blood test and physical examination are combined: both outcomes may be underdetermined, but the overall result is more robust.

The example of acute myocardial infarction may illustrate the principles described above. I make use of the description of the disease as given by Rubin and Farber (1999, p. 560). One of the early symptoms of acute myocardial infarction is severe substernal pain, which may simulate indigestion and may be preceded by symptoms of angina. However, one fourth to one half of all non-fatal cases occur without symptoms. In addition, the symptoms may also occur in cases of sliding hiatal hernia, an innocent defect in the diaphragm around the esophagus. The probability of a false positive or false negative diagnosis using these symptoms is relatively high. The diagnosis of acute myocardial infarction may be confirmed (this is the word used by Rubin and Farber) by either electrocardiography or the determination of certain enzymes and proteins which may appear in the blood. Without going into the details of the methods used for both there is no doubt that they are based on different theoretical and methodological principles: electrophysiology and biochemistry respectively. A fourth method, not mentioned by Rubin and Farber, is of course the finding at autopsy of myocardial necrosis (heart muscle death) that can be related to the time of the onset of the symptoms, but this can not be used in a living patient. It may be used to confirm the diagnosis in a deductive way (a test of the hypothesis).

In essence, the confirmation of the symptoms of chest pain by either electrocardiography or blood proteins is inductive confirmation and is a process of multiple derivability. The special feature of this type of confirmation is the more or less prescribed order of the test events; physical examination or any other test with a greater probability of false positive results are followed by tests with an overall higher sensitivity and specificity. This sequence is dictated by such conditions as the place where the test is done (at home at the bedside or in the hospital) and the costs and the simplicity of the test.

The case of the myocardial infarction is interesting for another reason. In general, a patient with a complaint of angina pectoris (substernal pain) will be sent by his family doctor to a see a cardiologist. The latter will make the laboratory determine enzyme levels in the blood and an electrocardiogram of the alleged patient will be made during physical exercise. When both tests are negative the diagnosis of heart problems will be rejected and the patient will be advised to consult a specialist in internal medicine, because sliding hiatic hernia has become a more probable diagnosis. In this example a diagnostic hypothesis degenerates because of lack of multiple derivability. This type of (lack of) inference is related to Thagard's criteria (1978) for inference to the best explanation: consilience of inductions, or multiple derivability, is one of these criteria. This seems to warrant the conclusion that lack of multiple derivability is lethal to a theory, but I shall argue below that this is not necessarily the case.

5.3.6 Some Thoughts About Reproducing and Replicating Tests, or How does a Difference Make a Difference?

Why do we not accept that mere repetition or exact replication of observations is sufficient evidence for a local theory (making us go to multiple derivability instead)?

The first answer to this question is an intuitive one: we may have made an error in the observation and by repeating the observation the error may be repeated as well. Several of the foregoing cases point in that direction. We may know we have been hallucinating and therefore we cannot trust the observations we have made. We have made a diagnosis but, due to low sensitivity and specificity of the diagnostic test, it may be wrong; we have to improve our diagnosis and repeating the same test will not help us in this case. We find fingerprints of the suspect in the room where the crime has been committed but finding more of the same fingerprints will not help us in improving the evidence of the guilt of the suspect.

Nevertheless, repeating the observation is a necessary step in the process of theory-making when the observation is made for the first time. The case of the invasion experiment is a good illustration. Suppose we do an invasion experiment with a new strain of bacteria, we count the colonies of bacteria on the agar plate and find surprising results (we neglect here why it is surprising). Three responses to this result are possible: cherish the result as the start of a new and promising research project, discard the result because it is surprising and does not conform to the core of our knowledge, or repeat the experiment to see whether the result is not in some way a coincidence; when the result will turn up again and again each time the experiment is repeated the possibility of a surprising novelty may be taken more seriously: apparently the new strain of bacteria is a source of new insight in the invasion process. The third response to the unexpected result is considered to be the choice that scientists usually make. In addition, repeating an experiment is essential in the process of acquiring practical skills: a student who has to learn to do invasion experiments has to show his skill by repeating the experiment and obtaining the same result as his predecessors.

The second answer to the question of why we do not accept that repetition of observations is sufficient evidence for a theory is that we may make a mistake in our inference. This is a problem of underdetermination and can be illustrated by using the invasion experiment again. Because of the way the experiment is performed it can be concluded that the colonies we count on the agar plate are representing the bacteria that were protected against the antibiotics because they were intracellular. However, at least two alternative explanations are available. (1) that the bacteria adhering to the cells may have induced the cells to produce an enzyme in the immediate vicinity of the bacteria that may protect the latter against the antibiotic, e.g. by degrading it; (2) that the bacteria have migrated between two cells to the underside of the cell layer where, after closure of the space between the two cells, the antibiotics cannot reach them. These extracellular bacteria may then be counted as colonies on the agar plate. Thus, a new and different experiment, that does not make use of killing extracellular bacteria by antibiotics, is necessary.[5]

[5] With regard to the question how many independent derivations are needed, it may be remarked that underdetermination may be infinite but in the practice of laboratory experiments or diagnostic tests two or three are thought to give sufficient arguments, because other alternative hypotheses, which are found too light when weighing on the basis of plausibility are discarded without testing.

How different should this new experiment be? Collins has shed light on this problem in his *Changing order*, where he discussed the principle of replication with a difference (Collins 1992, see especially Chapter 2). He opens the paragraph dealing with this principle with the statement that "[f]or an experiment to be a test of a previous result it must be neither exactly the same or be too different." He then goes on to make the difference between a first and a second experiment greater step by step: do it the next day yourself again, let the experiment be done by a colleague, or by a colleague using a similar apparatus built by someone else, or rather using an apparatus built for different purposes, and so on, until he arrives at a gipsy, reading the same result from the entrails of a goat. He then goes back again, from the gipsy and using old-fashioned equipment via a high school student to a first-rate physicist, all giving different confirmation depending on what apparatus they used. He then concludes that somewhere, i.e. at a certain degree of difference between the exact replication of the first experiment and the gipsy with his goat, an optimum in confirmatory power might be found; Collins shows a curve for this relation between experimental difference and degree of confirmation (p. 37). Beyond this optimum, where the difference gets gipsy-like dimensions, confirmatory power decreases gradually again. More importantly, when such differences present themselves pseudo-science may have cropped up.

My objection to this description is that confirmatory power does not increase or decrease gradually but that large parts of it turn up or disappear *en bloc*. This can be made clear by using the hierarchical model of theory making given before. In this hierarchy confirmation may take place at each of the levels of the hierarchy. The first type of confirmation is confirmation of the observation when it is made for the first time and is new and surprising. In this case it is essential that the second experiment is as similar to the first experiment as is possible, to exclude the possibility that the observation is a coincidental artefact of the experimental manipulation of the objects of study. In a second stage the experiment may be repeated by varying the methods, but not so varying as to change the bulk of background theories of the experiment. This variation is directed at confirming the observation, to ensure that it is generalisable and that it can be made under different circumstances. The steps described by Collins, such as doing it with an apparatus built for other purposes may sometimes be part of this strategy. Another example is the application of different procedures for fixation of tissues for their examination with the electron microscope (Nederbragt 2003; Culp 1994). The third type of confirmation is not directed at testing the observation but at testing reliability of the inference from the observation to the local theory. According to the strategy of multiple derivability which is applicable here, this experiment should be independent of the first with regard to methodological and theoretical principles. Studying the entrails of a goat fulfills this criterion of methodological and theoretical differences (it may therefore be seen as triangulation), but here we reject immediately and with force the whole body of background knowledge of this method, making Collins' curve falling abruptly back to the level of zero confirmation instead of sloping gradually downward.

It appears that repeating experiments contributes to making theories, either at the level of the observation or at the level of the inference from the observation to the theory. However, my hierarchical approach does not help in solving Collins' complaint (p. 38) about the difficulties with judgements of similarity or difference: experimenters may regard the difference of a new method unsatisfactory for the strategies of multiple derivability because they do not accept the background theory of that method.

The third answer to the question of why we do not accept that repetition of observations is sufficient evidence for a theory, turns up when we consider the theory under study as local and we want to reach the situation in which the theory, based on this experiment, has a more general meaning, not restricted to the experiment performed by this particular experimenter, in this particular laboratory. Radder (2003) has discussed this aspect and called this type of support for theories replication, "an experimental result [being] *replicable*, if it has been successfully applied to a certain domain and if it *might* be realized in one or more new domains". Again multiple derivability is at work here, although in a somewhat different form compared to that discussed above under the heading of the second question. In the former case difference and independence of methods was chosen to deal with the problem of underdetermination of the local theory, and the methods were different in background theories. In the latter case, difference and independence were chosen to make a theory leave its context, but not its background theory, since the new experiments accept the theory, but try to find robustness for this theory as it is, including its possible underdetermination, in a process of triangulation as I described it above.

With regard to the question of how much difference and independence we need for making theories more robust by multiple derivability we now may come to some conclusions.

1. There exists a hierarchy of derivabilities. Several cases point to this. We may be able to become more certain of the capability of bacteria to invade cultured cells but the question whether bacteria may be present inside the milk epithelial cells of living cows needs triangulation. This is, as it were, hierarchically one step higher and the in vitro experiments of invasion together form only one of several derivabilities on this new level. For the answer to the question whether invasion of bacteria into the udder cells of cows contributes to the pathogenesis of mastitis another set of derivabilities is required, again one step higher in the hierarchy.
2. Multiple derivability as a strategy only works in the context in which it is supposed to make use of two independent methods. We may question the context and investigate the background knowledge of that context. In the invasion experiment a local theory was addressed; multiple derivability could only be used for this local theory, i.e. the theory that bacteria may have invaded the cells cultured in vitro. In the epidemiology case no independent derivations ware available that helped us to conclude with more certainty based on database analysis alone that the alleged risk factor was the cause of the disease in question; triangulation was the convincing strategy here. In the diagnosis a theory is made about the disease of a patient; multiple derivability helps us to improve that theory, but it still is a theory of that patient alone.

3. What the case of the crime shows us is that background knowledge may help us to decide on the independence of methods. As soon as it is concluded that the background knowledge of one piece of evidence has an overlap with another piece of knowledge independence is reduced, and the theory may be overestimated. In the case of the invasion experiment it appeared that underdetermination of the theories that result from both methods has a role to play. When one of the methods can or has been used to validate the alternative method (or parts of it) then it is of no (or less) concern whether one or the other is used and the theory in question is not improved. Thus, absence of overlap in background knowledge of both methods and no validation of one method by the other seem to be crucial to decide on independence.
4. We may consider the repetition of an observation as an effort to obtain robustness through more derivabilities. When the observation is done by an independent observer we may conclude that she and I have seen a real pencil on the table although investigation of the mental state of both of us should elucidate whether hallucination might have played a role. In scientific practice, however, repeating an observation is used to test the observer (see Star 1986), which is not multiple derivability but validation, such as is done in the validation of laboratory methods by sending standard samples to different institutes that use the same machine and protocol for the analysis of patient samples.
5. Robustness may come gradually and in degrees. According to Wimsatt (1981) "different degrees of robustness may confer different degrees of stability". The case of the criteria for causation may be an example. The criteria 1–6 are all defined in statistical terms, to be applied to data sets of population parameters. Each time one of the criteria is fulfilled, the trust in the hypothesis that the factor under study causally contributes to the disease incidence increases, and this is a gradual process. However, the trust increases considerably when criterion 7 of the experimental approach is met. Reversely, when criterion 7 cannot be shown to be applicable the trust based on 1–6 may collapse. Perhaps we should not trust theories that we gradually arrived at by derivabilities that are not independent; these may be biased, as Wimsatt (1981) has shown for theories of group selection in evolutionary biology. It may be difficult to decide whether or not derivations are sufficiently independent; in daily practice scientists discuss experimental methods and their results but in their evaluations independence is, I think, a hidden criterion, applied intuitively.

5.4 Stabilizing Knowledge Without Robustness

In the preceding paragraph I introduced the question whether local theories that were not arrived at by multiple derivability or other strategies of robustness might degenerate by lack of stabilization. Here I will argue that this is not necessarily the case. In doing so I have to make a distinction between reliability and stabilization of theories. Reliability is the outcome of epistemological strategies alone whereas stabilization, e.g. within a scientific community, is a mixture of epistemological reliability and social acceptance, in which negotiations between actors involved in the theory have played a role.

In the earlier paper I referred to and in which I discussed strategies for obtaining more certainty for theories (Nederbragt 2003) I proposed that three ways of anchoring of knowledge may be distinguished, leading to its stabilization. I suggested that these ways of anchoring may reduce the impact of the problem of the experimenter's regress. The first is epistemological anchoring, for instance by multiple derivability or triangulation, leading to more robustness of the theory. The second is historical anchoring, i.e. the theory may be rooted in a much older but stable theory that, in a coherent way, became transformed into a different but more recent theory that derives it stable character from the older theory; my example was the older theory of phagocytosis of bacteria by cells giving rise to a more recent theory of invasion of bacteria into cells which was later shown to be a different process. The third anchoring was social anchoring which I will discuss in more detail here.

Although Rheinberger (1997, p. 91) thinks that scientists do not accept sociological solutions for problems such as the experimenter's regress and that they try to stabilize their facts by improving their experimental systems, I will nevertheless argue that social processes do play a role in stabilization of theories when these are, at first sight, not sufficiently reliable in an epistemological sense.[6]

Social anchoring is the process in which the stabilization of a piece of knowledge is based on the mutual interaction of theory, method (machine, protocol), scientist and scientific community. In my view social anchoring may also be seen as a strategy that may be used to overcome the problem that turns up when theories cannot be stabilized by robustness. In cases in which only one derivation to a theory is possible the theory may degenerate by lack of support from a second derivation (see my discussion of the bacterial mesosome in Nederbragt 2003) or it may be stabilized by social interactions. An example of the latter stabilization is obtained from the discipline of immunohistochemistry (for a detailed description of this case, see Nederbragt 2010).

The case is an experiment in which tissue samples of tumours of dogs are fixated, prepared and sliced in such a way that a very thin section of the tissue, glued onto glass, can be studied with the light microscope. The aim of the study was to stain a special protein in the tissues, tenascin-C in this case, in order to make it possible to investigate the pattern of its distribution in normal mammary glands and malignant mammary tumours and to find a possible correlation of this tenascin-C pattern with the benign or malignant behaviour of the tumour, i.e. its growth, infiltration and metastasis into other tissues. Tenascin-C is a protein that may accumulate outside cells of normal tissues and tumours and may be present in microscopically visible

[6] Epistemological arguments lead to more reliability of a theory, whereas social arguments lead to stability of that theory. Theories may be reliable but not stable because they do not fit in a reigning paradigm (see Gillies (2005) for an analysis of the case of Semmelweis's childbed fever). Theories may be stable, although they are epistemologically unreliable. Historical anchoring may contribute to both.

fibers and bands. In this study the findings of a tenascin-C pattern could be compared to that of human tissues and tumours, which were well studied at that time; in dogs this had never been done before.

Three important stabilization steps may be distinguished in this research project: the stabilization of the method as a personal achievement of the investigator, the stabilization of the pattern (or patterns) as an achievement of the investigator and the group of which he is member, and the stabilization of this pattern as an accepted theory in the wider science community.

The method for detection of the protein is based on applying an antibody against tenascin-C, marked with a chemical stain, on the tissue slices that reacts specifically with tenascin-C where it is present in the tissue and that consequently is visible in the light microscope. There is no alternative method available for detecting the protein that is theoretically independent. A biochemical method for measuring tenascin-C gives loss of the tissue structure; with electron microscopy a comparable antibody has to be used, albeit with a different marker attached to it, and the high magnification prevents getting an overview of the complete tissue. Thus, when we consider the protocol for tenascin-C staining of canine tissues as some local hypothesis, this is then based on a single derivation; multiple derivability seems impossible here.

The same holds for the evaluation of all the tissue slices to detect and construct one or more patterns of tenascin-C distribution: it can only be performed by microscopic investigation of the tissues and the patterns are constructed in a process of recognition of the individual tenascin-C distribution in each of the slides and their subsequent grouping by the investigator. Multiple derivability does not play a role.

It may be argued that both the method of detection of tenascin-C and the pattern of its distribution are strongly rooted in the background knowledge available from studies with human tumours. This is certainly the case. However, it cannot be decided a priori that tenascin-C in dogs and humans have the same biochemical features and have similar distributions in both species. Besides, although mammary tumours of dogs and women share many characteristics they are different in many other respects. In addition, the first efforts to detect tenascin-C in human tissues were part of a triangulation process of immunohistochemistry and biochemistry, but this type of triangulation was not available for canine tenascin-C. At most, the local theory of canine tenascin-C derived some reliability from the analogy with the human protein.[7]

How then became method and pattern stabilized knowledge? In my opinion this is achieved by a process which is described by Fleck (1979) for the Wassermann test as a tool for diagnosing syphilis; it can be summarized in Flecks words as follows: "[The] thought collective ... standardized the technical process with genuinely social methods, at least by and large, through conferences, the press, ordinances and legislative measures." Knowledge becomes stabilized in the social interaction

[7] The possibility to anchor a local theory into the background knowledge of a related field because of analogy may be an important epistemological argument in daily practice of research (see Thagard 1978). In the discussion of causal criteria in the former section the plausibility of criterion 8 seemed to play a comparable role.

between investigator, his research group, the science community and society. A similar but smaller-scale and more local process took place during standardizing immunohistochemical tenascin-C detection and the study of its distribution in tissues. Initially the development of a proper protocol for tenascin-C staining of tissue slices was the "private" work of a single biomedical scientist. He checked all the steps of the protocol by evaluating the staining of the tissues with the microscope. By comparing several different slides he became aware of the repeatability of the protocol and therefore he could start with the search for a pattern of distribution of the protein. Arranging all the individual different distribution findings into groups of different types of distribution and constructing a pattern from this was also a new step in stabilizing the protocol: these patterns could be found with this protocol as developed by this investigator; protocol, pattern and investigator formed as it were a coherent unit.

In the next step the pattern as an interpretation of all the findings of the different slides had to become a theory that had to be made part of the set of theories of the research group. The investigator had to discuss the outcomes of representative individual slides with other members of this group, including his supervisor, and to defend his arguments and adapting his groupings when he could not convince his colleagues. When he finally succeeded in proposing a pattern of tenascin-C distribution that was acceptable for the whole group the knowledge of this pattern was incorporated into the "group knowledge", thereby stabilizing the theory of this pattern and further stabilizing the knowledge of the protocol. The knowledge that was initially only knowledge of a single investigator was now knowledge of the research group.

Similarly, by publication of this pattern, and its correlation to the knowledge of the clinical and morphological features of canine mammary tumours, the theory, including pattern and protocol, became stabilized knowledge in the science community in which the research group participates.

It cannot be denied that a certain degree of epistemological stabilization has taken place too, although it is difficult to decide in what degree; my suggestion of analogy and plausibility may be relevant here. But contrary to what happens when robustness is obtained by multiple derivability, in this case a considerable proportion of stabilization is obtained by social interaction. It may be an interesting study to investigate why some theories degenerate and others survive, although both groups of theories may have been anchored initially by social interactions, and whether the survival of robust, more epistemologically stabilized theories, is better than those that have been socially stabilized.

5.5 Conclusions

I started with discussing the characteristics of consilience of inductions, triangulation and multiple derivability and I concluded that there is a certain fluidity in the definitions of each; however in the way the concepts are applied there seem to be sufficient differences to make it possible to consider them as three different strategies, each having a place in the process of the development and analysis of theories. Consilience of inductions does not seem to be important as a strategy of daily scientific practice since, because of its emphasis on the coincidence and surprise, it seems to be connected to deductive reasoning after theories have become made reliable or stable. Triangulation and multiple derivability are strategies of practice, the former aiming at making better theories by the convergence of more local theories from various disciplines or fields, the latter aiming at better theories by the convergence of independent (experimental) derivations from different methods.

Next I investigated the problem of independence of derivations. Independence is necessary because, when derivations are somehow dependent on each other, they may lead to a theory that is overestimated and may reach a state of reliability that, on closer inspection, appeared to have been an illusion. I have argued that, to learn more about the independence of derivations, investigations of the degree of overlap in background knowledge on which they are based is necessary and that it is also important to find out whether the methods for the derivations are not validated by each other. The case of causal claims in epidemiological research, analyzed in this section, also demonstrated that when derivations are insufficiently independent triangulation may help to establish the claim with more certainty.

In the subsequent discussion about why experiments should be repeated and how different they should be it became clear that a hierarchy of various types of derivations may be discerned in scientific practice, each representing strategies for problems of theory improvement at its specific level. The strategy of triangulation fits nicely into such a hierarchical structure.

Theories may be reliable because of epistemological justification but in scientific practice something more is needed to give it stability in a scientific or social environment. A typical case is analyzed shortly in the third section. The social factors which are addressed here are derived from the interaction between protocol, scientist, research group and science community. The local theory, developed by the scientist interacting with the protocol, cannot gain sufficient robustness with the strategies described above: multiple derivability could not be applied because of lack of independent methods, nor could the theory be triangulated on theories of another discipline. Only analogy with theories based on background knowledge of other disciplines could be invoked. I argued that a subordinate theory at one level may become stabilized by its fitting into a more overarching theory at a higher level. In this process from one level to the next, a local theory may transform into a theory with a wider scope, although the contents may not change.

Both robustness strategies to reach reliability and social interactions to reach stability seem to be hierarchically structured. It may be interesting to investigate to what extent these structures are related to each other.

References

Collins, H.M. 1992. *Changing Order. Replication and Induction in Scientific Practice*. 2nd ed. Chicago: Chicago University Press.

Culp, S. 1994. "Defending Robustness: The Bacterial Mesosome as a Testcase." In *PSA-1994*, vol. 1, edited by D. Hull, R. Forbes, and R. Burian, 46–57. East Lansing, MI: Philosophy of Science Association.

Döpfer, D., R.A. Almeida, T.J.G.M. Lam, H. Nederbragt, S.P. Oliver, and W. Gaastra 2000. "Adhesion and Invasion of *Escherichia coli* from Single and Recurrent Clinical Cases of Bovine Mastitis In Vitro." *Veterinary Microbiology* 74:331–43.

Evans, A.S. 1993. *Causation and Disease. A Chronological Journey*. New York: Plenum.

Fisch, M. 1985. "Whewell's Consilience of Inductions – An Evaluation." *Philosophy of Science* 52:239–55.

Fleck, L. 1979. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Gaudillière, J.-P. 1994. "Wie Man Modelle für Krebsentstehung Konstruiert. Viren und Transfektion am (US) National Cancer Institute." In *Objekte, Differenzen und Konjunkturen. Experimentalsysteme in historischen Kontext*, edited by M. Hagner, H.-J. Rheinberger, and B. Wahrig-Schmidt, 233–57. Berlin: Akademie Verlag.

Gillies, D. 2005. "Hempelian and Kuhnian Approaches in the Philosophy of Medicine: The Semmelweis Case." *Studies in the History and Philosophy of Biology and Biomedical Sciences* 36:159–81.

Hale, T.L., and P.F. Bonventre. 1979. "Shigella Infection of Henle Intestinal Epithelial Cells: Role of the Bacterium." *Infection and Immunity* 24:879–86.

Hale, T.L., R.E. Morris, and P.F. Bonventre. 1979. "Shigella Infection of Henle Intestinal Epithelial Cells: Role of the Host Cell." *Infection and Immunity* 24:887–94.

Laudan, L. 1981. "William Whewell on the Consilience of Inductions." In *Science and Hypothesis. Historical Essays on Scientific Methodology*, edited by L. Laudan, 163–80. Dordrecht: D. Reidel.

Nederbragt, H. 2003. "Strategies to Improve the Reliability of a Theory: The Experiment of Bacterial Invasion into Cultured Epithelial Cells." *Studies in the History and Philosophy of Biological and Biomedical Sciences* 34:593–614.

Nederbragt, H. 2010. "Protocol, Pattern and Paper: Interactive Stabilization of Immunohistochemical Knowledge". *Studies in the History and Philosophy of Biological and Biomedical Sciences* 41:386–95.

Nickles, T. 1989. "Justification and Experiment." In *The Uses of Experiment. Studies in the Natural Sciences*, edited by D. Gooding, T. Pinch, and S. Schaffer, 299–333. Cambridge: Cambridge University Press.

Radder, H. 2003. "Technology and Theory in Experimental Science." In *The Philosophy of Scientific Experimentation*, edited by H. Radder, 152–73. Pittsburg, PA: The University of Pittsburgh Press.

Rheinberger, H.-J. 1997. *Towards a History of Epistemic Things. Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Risjord, M. 2001. "Methodological Triangulation in Nursing Research." *Philosophy of the Social Sciences* 31:40–59.

Rubin, E., and J.L. Farber 1999. *Pathology*. 3rd ed. Philadelphia, PA: Lippincott-Raven.

Star, S.L. 1986. "Triangulating Clinical and Basic Research: British Localizationists, 1870–1906." *History of Science* 24:29–48.

Thagard, P.R. 1978. "The Best Explanation: Criteria for Theory Choice." *Journal of Philosophy* 75:76–92.

Whewell, W. 1847. *The Philosophy of the Inductive Sciences, Founded Upon Their History*, vol. 2. London: John W. Parker.

Wimsatt, W.C. 1981. "Robustness, Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins. San Francisco, CA: Jossey-Bass Publishers.

Chapter 6
Robustness of an Experimental Result: The Example of the Tests of Bell's Inequalities

Catherine Dufour†

6.1 Introduction

The general aim of this chapter is to discuss the robustness of physics experiments, more precisely, the robustness of the result of experimental tests of a given theoretical question. The chosen example is the result of the experimental tests of Bell's inequalities. These tests provide a quantitative criterion for discriminating experimentally between local hidden variable theories (LHVT) and standard quantum mechanics (SQM).

This introduction begins with a short historical background and then presents the aim of the chapter and its contents.

† Deceased.

Note by Léna Soler. Cathy Dufour prematurely passed away in March 2011 at the age of 46. She first trained as a physicist and worked as an experimenter in a condensed matter physics laboratory (the Institut Jean Lamour in Nancy) for more than 20 years. At one point in her career, she realized that she had to face multiple important changes in her scientific domain, and she felt the need to better understand her experimental practice from a philosophical point of view. To that end she began to study philosophy of science at the University of Nancy, which is where I first met her as a teacher. Some years later, in 2007, when I created the PratiScienS research group in Nancy (where PratiScienS stands for: "Rethinking science from the standpoint of scientific practices"), she became an active member of this group. She asked us to organize an ethnographic study in her scientific laboratory, and continuously pressed the members of the research group to deepen and improve our analyses of what happened there. She was a pivotal contributor of PratiScienS, especially with respect to the robustness theme. She was about to begin working on a PhD devoted to the robustness problem when she suddenly died.

L. Soler (✉)
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France
e-mail: l_soler@club-internet.fr

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_6,

6.1.1 Historical Background

To improve our understanding of nature, a number of scientists think that it is crucial to answer the following questions:

> Is nature intrinsically probabilistic (the SQM point of view) *or* do quantum mechanical properties derive from our ignorance of some hidden parameters and, therefore, an underlying deterministic theory is conceivable (the LHVT point of view)?

The first point of view is largely accepted by the physics community today. (However, its assimilation by the wider culture is very limited.) In fact, given that a large number of theoretical predictions deriving from this theory have been confirmed by very accurate experiments, SQM is one of the pillars of modern physics.

The second point of view was originally suggested in 1935 by Einstein, Podolsky and Rosen (EPR).[1] They claimed that SQM is not complete in the sense that not every element of reality has a counterpart in the theory. Their argument employed a by now well-known "gedanken experiment", the so-called EPR experiment.

In 1964, Bell[2] proved a famous theorem that made it possible to test LHVT against SQM quantitatively. According to this theorem, any realistic LHVT must satisfy certain inequalities (the so-called Bell inequalities) that can be violated in SQM, thus allowing an experimental test of the validity of SQM versus LHVT.

Since the end of the 1960s, several generations of experiments have been performed in order to try to reproduce an EPR type "gedanken experiment" and thus to test Bell's inequalities.

The total number of experiments that have been carried out by now is very large. All but one of these tests are in substantial agreement with SQM and strongly disfavour LHVT. In the following, this statement will be considered as *the* result of the experimental tests of Bell's inequalities. That is, we have good agreement with the SQM predictions and disagreement with the LHVT predictions, or, for short, experimental confirmation of SQM and refutation of LHVT.

6.1.2 Aim of the Chapter

Following Soler,[3] the different experimental tests of Bell's inequalities can be designated "experimental lines of argument", lines that correspond to the multiple derivations invoked in Wimsatt's seminal 1981 paper on robustness.[4]

In the following, we will address the question of robustness at two different levels:

---

[1] Einstein et al. (1935).

[2] Bell (1964).

[3] Soler (Chapter 10).

[4] Wimsatt (1981), Reprinted in this volume, chapter 2.

(a) We will first consider the whole set of experiments that have been performed since the 1970s and their respective outcomes. We shall discuss independence of this large number of experiments performed during the last 40 years and the invariance of their outcomes. We will argue that, at a first glance, one can assert that the result of the experimental tests of Bell's inequalities is robust if one follows the statement of Wimsatt[5] (1981): the robustness of a result is characterized by its invariance with respect to a great number of independent derivations.
(b) We will then consider the individual experimental derivations and discuss their internal solidity. From an experimental point of view, the individual derivations can be considered accurate and reproducible. So one can be tempted to conclude that they are solid. However, the important point is that these derivations are not ideal. They do not exactly reproduce the "gedanken experiment" used to derive the Bell's theorem, even if, since the end of the 1960s, they have evolved closer and closer to the ideal EPR scheme. Despite these continuous improvements, all of them, taken individually, present conceptual failings or biases, or (to use the vocabulary of the scientists working in the field) "loopholes." Two kinds of problems are mainly emphasized:

  - In all the experiments, an additional assumption is necessary, due to the fact that a part of the experimental set-up is not 100% efficient. This leads to "the detection loophole."
  - The experiments do not fulfil completely one of the requirement of the theorem, the locality condition. This leads to "the locality loophole."

Consequently, strictly speaking (especially for practitioners who take these loopholes seriously), one cannot conclude that experimental tests have ruled out the LHVT. The occurrence of these loopholes will lead us to question the solidity of all the individual derivations and consequently the robustness (in Wimsatt's sense) of the result of the tests. The main question can be summarized as follows: Do the loopholes constitute a failure to achieve robustness?

6.1.3 Contents

The next section of this chapter will present the ideal entanglement experiment, Bell's theorem, and its experimental tests. Then, in a third section, we shall consider the whole set of experiments that have been performed since the 1970s with regard to the robustness of their overall result (in Wimsatt's sense). Section 6.4 will question the solidity of the individual derivations and investigate the influence of the loopholes on robustness. Finally, in a fifth section, we shall attempt to achieve a deeper understanding of the robust character of an experimental result by discussing

[5] Wimsatt (1981).

the following questions: Have some of the aforementioned experiments been considered more crucial than others? Are the two kinds of loopholes equally important? Are they both crucial? Is an ultimate experiment that closes both loopholes simultaneously desirable and/or necessary to conclude that the result favouring SQM is robust? Or, are a couple of experiments, each one closing a given loophole, enough?

6.2 Bell's Theorem and Its Experimental Tests

6.2.1 *The Ideal Entanglement Experiment ("Gedanken Experiment")*

An experimental test of the Bell's inequalities requires the use of entangled systems. In fact, as pointed out by EPR, for certain quantum states (called entangled states), SQM predicts a strong correlation between distant measurements.

An entangled system is constituted of two (or more) particles previously in interaction for a finite time and now separated and moving in different directions. Before a measurement, the values of a given observable are unknown. The measurement of the observable of one particle instantaneously fixes the value of the correlated observable of the second particle, independently of the distance between the particles. The optical variant of David Bohm's version of the EPR "gedanken experiment" is presented in Fig. 6.1.

A source S emits a pair of photons with different frequencies $\nu 1$ and $\nu 2$, counterpropagating along Oz. Suppose that the polarization part of the state vector describing the pair is: $\psi\,(1,2) = \frac{1}{\sqrt{2}}\left[Ixx > +Iyy >\right]$ where x and y are linear polarization states. This means that if the polarisation of the first particle is measured along x, then the polarisation of the second one will be also along x.

This state is noteworthy in that it cannot be factored into a product of two states each associated with a different one of the two photons; so we cannot ascribe any well-defined state to each photon. In particular, we cannot assign any polarization to each photon. Such a state, describing a system of several objects that can only be thought of globally, is an entangled state. Linear polarization measurements are now performed on the two photons, with analysers I and II. Analyser I, in orientation **a**, is followed by two detectors, giving results $A = +1$ or $A = -1$, corresponding to a

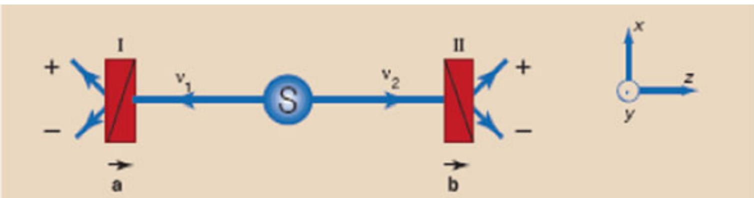

Fig. 6.1 EPR experiment with photons. S is the photon source, I and II are analysers used to perform polarization measurements, and **a** and **b** are the orientations of each polarizer

linear polarization found parallel or perpendicular to **a.** Analyser II, in orientation **b**, acts similarly. The outcome of the measurement is now $B = 1$.

In summary, the global state of an entangled quantum state of two particles is perfectly defined, whereas the states of the separate particles remain totally undefined. The information contained in an entangled state is all about the *correlation* between the two particles. Nothing is said (can yet be known) about the states of the individual particles.

6.2.2 Bell's Theorem

Introducing a locality condition, Bell's theorem states that:

1) LHVTs are constrained by Bell's inequalities.
2) Certain predictions of SQM violate Bell's inequalities and therefore refute the claim that SQM is compatible with LHVT.

On the one hand, Bell's inequalities are inequalities between correlation coefficients. They are valid if we assume the occurrence of local hidden variables. On the other hand, the correlation coefficients can be easily calculated using the SQM formalism. In some situations, the result of the calculations using SQM violates Bell's inequalities. Up to the present there has been no criticism of the derivation of Bell's theorem. In fact, according to several authors,[6] Bell has translated the theoretical and philosophical question about the nature of reality into a problem for experimental physicists.

The fundamental assumptions absolutely necessary to obtain Bell's inequalities and consequently a conflict with SQM are:

- the existence of supplementary parameters (hidden variables) that yield an account of the correlation at a distance
- the locality assumption: (a) the outcome of a measurement A with I does not depend on the orientation **b** of II and vice-versa; (b) the way the pair is emitted by the source does not depend on **a** and **b**.

This is why one often claims that SQM conflicts with local realism.

It is easy to generalize the formalism to Stochastic Supplementary Parameter Theories: the deterministic measurement functions of the polarisations are then replaced by probabilistic functions. Bell's inequalities still hold, conflict does not disappear; and so the deterministic character of the formalism is not the reason for the conflict.

[6] Percival (2000a), Aspect (2004). Abner Shimony forged the now famous expression "experimental metaphysics" in order to characterize the situation.

For a more detailed presentation of Bell's theorem and further discussions, the author invites the reader to refer to the papers of Aspect[7] and Berche et al.[8]

6.2.3 Experimental Tests of Bell's Inequalities

The aim of the experimental tests is to reproduce an EPR type "gedanken experiment", to measure the correlation coefficients, and to compare the outcomes with the prediction of SQM. Many experiments have been performed in the past 40 years or so (see Fig. 6.2).

Usually, one considers that there are two groups of experiments: experiments performed with photons and experiments performed with massive particles. The first group is then divided in two subgroups: the one using γ photons and the one using optical photons. The optical photons group again can be divided in two subgroups:

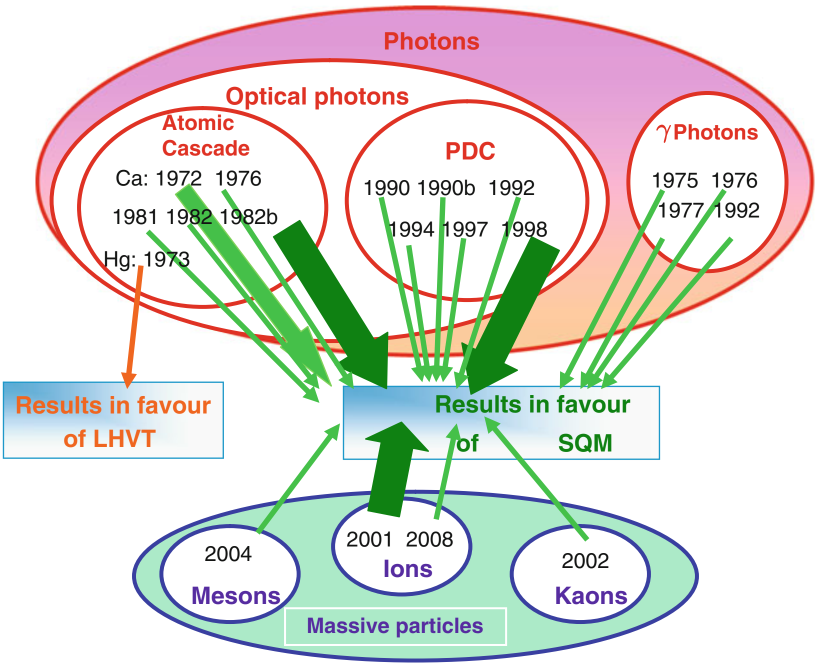

Fig. 6.2 Classification of a large number of experiments performed in order to test Bell's inequalities (the numbers are the years of publication). The *ellipses* contain groups or sub-groups of experiments using the same particles, or the same way to produce the particles. The *arrows* show the kind of result obtained (either in favour of SQM or in favour of LHVT). The widths of the *arrows* represent the importance of the paper in question

[7] Aspect (2004).

[8] Berche et al. (2006).

- experiments performed with optical photons produced using atomic cascade decay, including the first generation experiments and the experiments of the Orsay group that led to substantial improvement, mainly an improvement toward meeting the locality condition.[9]
- experiments with pairs of photons produced by parametric down conversion.[10]

Figure 6.2 presents a sketch of a large portion of the experimental tests of Bell's inequalities performed between 1972 (see Freedman and Clauser 1972 for the first experiment performed) and 2008. The diagram will provide an iconic point of departure with respect to further discussion of independent derivations, of the invariance of the result and of crucial experiments. For a more detailed description of the experimental tests of Bell's inequalities, the reader is invited to refer Genovese's paper.[11]

6.3 Invariance of the Outcomes of Bell's Inequalities Tests with Respect to a Great Number of Determinations

According to Wimsatt's 1981 paper, the robustness of a result is synonymous with invariance with respect to a sufficiently large number of independent derivations. Accordingly, we are now going successively to address the following questions: Do we have multiple means of determination in the case of Bell's inequalities? Are the various experimental tests independent? Are the multiple derived experimental outcomes invariant? These three questions can be summarized in one basic question: Are we in presence of a robust result in Wimsatt's sense?

6.3.1 Multiple Means of Determination

According to the above definition of robustness, the use of *multiple means of determination* is needed to triangulate on the existence and character of a result. Concerning the test of Bell's inequalities, we definitely have multiple means of determination, since dozens of experiments have been performed up to now. It is necessary to underscore that the list presented above (Fig. 6.2) is not exhaustive at all.

At this point a subsidiary question can be addressed. One can note that the number of experiments has greatly increased in recent years. Why has such a large number of experiments been performed over the last 40 years or so in order to

---

[9] This improvement has been achieved by changing the settings during the flight of the particles.

[10] The parametric down conversion (PDC) is a quantum effect and consists in a spontaneous decay, inside a non-linear crystal, of a photon from a pump beam (usually generated by a laser) into a couple of photons. This process obeys to energy conservation (phase matching laws).

[11] Genovese (2005).

test Bell's inequalities, and why are they still being performed now? Here are the answers.

(1) First of all, because the underlying question (local realism) appears as a very crucial question (to the physicists as well as to the philosophers of science); it has a large conceptual relevance.
(2) Moreover, because of the occurrence of loopholes in the experiments.

Other reasons could also be called for to explain the occurrence of such a large number of experiments testing Bell's inequalities. For instance:

- entanglement seems fascinating
- the experimental know-how concerning creation and manipulation of quantum systems has hugely increased during the last decade
- entanglement phenomena are closely linked to quantum computation. For example, many groups are involved in this research area because of possible applications.

6.3.2 Independence of the Various Experiments

According to Wimsatt we need to analyse a variety of (at least partially) independent measurement processes in order to draw conclusions regarding robustness. Are the experimental tests of Bell's inequalities independent? Following Soler's introduction of this book, we can distinguish content-independence and genetic-independence.

Let us consider first the content-independence. Most of the experimental tests performed are entanglement experiments. All of them involve particles and include their source, an analyser of a given observable, a detection system and additional experimental set-ups. So one can conclude that we are not in the presence of strong independence. However, we can state that there are groups of experiments that are independent, because all the above elements are different from one group of experiments to the other:

- The various experiments use different particles: photons (γ and visible photons) and massive particles (protons, mesons, ions, kaons);
- For a given particle, several means of production involving different physical phenomena have been used: for example optical photons are produced either by atomic cascade, or by parametric down conversion in non linear crystals;
- The experiments measure different observables: for example, for optical photons we obviously think of polarisation measurements, but also energy and time, position and momentum have also been measured;
- They use different types of detection.

With respect to genetic independence, we will only stress that the experiments have been performed by different teams of physicists all around the world, belonging to different communities (human independence[12]). However, in each group, some experiments are only partially content-independent. In fact, some experiments are taken to be an improvement of the previous one and some of them are performed by the same group of researchers, or by researchers connected to a previous group.

6.3.3 Invariance of the Result

All the outcomes of the experimental tests of Bell's inequalities, except for one, substantially agree with SQM and strongly disfavour LHVT. So in a sense, ignoring the unique exception, or, more exactly, dismissing it as unsound and physically insignificant (as most practitioners tend to do today), one can state that the outcomes of the multiple experiments lead to one and the same invariant result: 'confirmation of SQM and refutation of LHVT'. However, since the experiments are different, their outcomes are actually not identical, and according to me, it is more appropriate to talk about convergent, compatible outcomes rather than about an invariant result.

The exception (i.e. the test that favours LHVT) merits a special discussion. The philosophical point is that established theories (here SQM) are protected. This was mainly stressed by Lakatos.[13] A good example of this behaviour happened in the early atomic cascade test of Bell's inequalities. As already noted, a first experiment performed by Freedman and Clauser (1972) was in agreement with SQM, and a second experiment made by Holt and Pipkin disagreed with the SQM predictions, since it did not violate the inequality tested. The consequence is that the latter experimental results were never formally published and many people (including the authors) made a careful search for possible sources of errors. The Holt experiment has two differences from the Freeman one: the source (a Hg cascade instead of Ca) and the analysers (calcites polarisers instead of piles of plates polarisers). An experiment using an Hg cascade was performed later and came out in favour of SQM. However, an experiment using calcite polarisers has never been reproduced, despite the fact that there are arguments supporting the claim that these polarisers present interesting properties.[14]

[12] This is of course not a necessary condition to warrant genetic independence. But a complete discussion of genetic independence is beyond the scope of this paper.

[13] Lakatos (1980).

[14] Santos (2005).

6.3.4 A Possible Conclusion at This Point

If one holds with Wimsatt that invariance under multiple independent determinations (or, more exactly, convergence) is synonymous with robustness, then one can conclude that the result of the experimental tests of Bell's inequalities is robust. And, in fact, as we will comment later in Section 6.5, many scientists agree with this conclusion.

6.4 Loopholes: A Failure in Robustness

From an epistemological point of view, the solidity of each individual derivation is crucial for the robustness of the derived result. Following Soler,[15] we state that the examination of this solidity is an additional necessary condition for the discussion of the robustness of the result. Thus, complementary investigations are needed.

At a first glance, the accuracy and the reproducibility of each individual experimental test can be considered as a synonym for the solidity of the corresponding derivation. This point will be discussed in the next section. However, in my judgment accuracy and reproducibility are not the only conditions needed to achieve the solidity of the individual derivations: an additional condition is the ability of the actual experiment to instantiate the ideal EPR scheme. This point will be examined in Section 6.4.2.

6.4.1 Accuracy and Reproducibility of the Outcomes of the Various Independent Experimental Tests

In order to determine the accuracy and thus the reproducibility of a given experiment, one uses the standard deviation[16] σ in order to estimate the error on the set of experimental results, (in this case a result is a correlation coefficient or a combination of correlation coefficients). Then to estimate the accuracy (that is here in fact the ability to violate the Bell's inequalities), one compares the experimental value with the theoretical value. The difference is converted into standard deviation σ units. The larger σ is, the more convincing is the violation of equality, thus the greater is the inequality.

We can note that even in the 1970s the results were accurate (the violation was by 4 or 5σ) and that the accuracy has increased with time, mainly because of improvements of the technology. The best violation was obtained in 1998. The Weihs result violated the Bell's inequalities by 100σ. However, it is necessary to emphasize that, when there is a first attempt to close a loophole, a decrease in the accuracy

---

[15] Soler (Chapters 1 and 10).

[16] The standard deviation of a set of outcomes is defined as the root-mean-square deviation of the values from their mean, or as the square root of the variance.

of the result can be observed. Consider the third Aspect experiment,[17] for example. This experiment was an attempt to close the locality loophole, and it violated Bell's inequalities by 4σ; whereas the previous one, which did not close the locality loophole, violated them by 40σ.

Thus the statement of accuracy and reproducibility of the individual derivations can be accepted. This statement in turn supports the robustness of the outcomes of these derivations. Moreover, if one considers that publication in a prestigious journal is an indicator of solidity, one has to note that most of the papers concluding that SQM is valid have been published in top journals such as *Physical Review Letters* and *Nature*. All this strongly supports the robustness of the result of the experimental tests of Bell's inequalities.

6.4.2 The Loopholes

It is important to note a distinction between the accuracy and the reproducibility of the outcome of each individual experimental test on the one hand, and, on the other hand, the ability of this test to actually reproduce the "gedanken experiment". Concerning this last point, we state that real experiments, up to now, differ from the ideal experiment in several respects. There are two kinds of problems:

- a part of the experimental set-up is inefficient (or not 100% efficient)
- the experiment does not fulfil one of the requirements of Bell's theorem, for example the locality condition.

So a given experiment can fail as a genuine test of Bell's inequalities even though it is accurate and reproducible, and even though the experimentalists can be right to claim that their experiment is correct.

In fact, the performed experiments either do not test Bell's inequalities alone but instead test them plus additional assumptions, or else they test them with missing crucial assumptions. This is due to the fact that genuine Bell's inequalities are extremely difficult to test. In general, one considers that two main loopholes for the refutation of LHVT occur: the locality loophole and the detection loophole.

6.4.2.1 The Locality Loophole

One of the central premises of Bell's theorem – locality – is not easily experimentally verified. In fact, many of the test experiments have been static. For example, in the case of optical photon experiments before 1980, the positions of the polariser were fixed long before the detection events took place. Therefore, the experiment could not test locality, in the sense of relativistic causality.

[17] Aspect (1982).

To quote Bell: "the settings of the instruments are made sufficiently in advance to allow them to reach some mutual rapport by exchange of signals with velocity less or equal to that of light". Bell insisted upon the importance of experiments in which the settings are changed during the flight of the particles.[18]

The locality condition can be split in two conditions:

- the distant measurements on two subsystems must be space-like separated
- the choices of the quantities measured on the two separated subsystems must be made at random and must be space-like separated.

The second condition is obviously difficult to fulfil.

In 1982, in order to try to close the locality loophole, Aspect[19] and co-workers performed a new atomic cascade experiment where (in some sense) the polarisers' positions were chosen when the photons were already in flight. Aspect's experiment has for a long time been presented as the definitive refutation of local realism. However, it is not definitive – and for several reasons:

- lack of angular correlation because of using atomic cascade sources
- low efficiency of the detector for optical photons (see next section)
- optical switches driven by a periodic generator and thus not random.

In my view the condition of locality was largely satisfied for the first time only in 1998 by Weihs[20] and co-workers. In their experiment, the necessary space-like separation of the observations was achieved by:

- sufficient physical distance between the measurement stations (440 m)
- ultra-fast and random settings of the analysers
- completely independent data registration using local atomic clocks.

They observed a violation of the inequalities of 100 standard deviations.

This experiment has been considered as crucial (in Francis Bacon's sense) by a number of physicists. It became possible because of the development of new sources of correlated photons: in these sources a pair of red photons is produced by a parametric down conversion (PDC) of a UV photon in a nonlinear crystal. This experiment has been quoted as demonstrating weak non-locality or as a final re-enforcement of SQM, and has given to many the impression that the experimental

[18] Bell (1987).

[19] Aspect et al. (1982).

[20] Weihs et al. (1998).

quest is over.[21] Other experiments with great distances between the source and the measuring apparatus have been performed (4 km[22] and 10 km[23]).

However, recently (in 2008), the Gisin group[24] in Geneva emphasized that "none of the tests of Bell inequalities that have been performed so far involve space-like separated events". Why? Because, in these tests, no one has considered the so-called measurement problem. The common view among quantum opticians is that a quantum measurement is finished as soon as the photons are absorbed by the detector. The alternative view is that in quantum optics the measurement is finished once the alternative result would have led to the displacement of a sufficiently massive object. The authors assume a connection between quantum measurement and gravity: the measurement time is related to gravity-induced state reduction. This group performed an experiment with space-like separation large enough (18 km) to include a hypothetical delay of the quantum state reduction until a macroscopic mass had significantly moved.

6.4.2.2 The Detection Loophole

In most of the experiments an additional assumption is employed, due to the low total detection efficiency, namely, that the observed sample of particle pairs is a faithful subsample of the whole. This additional assumption is called the fair sampling assumption. Its use can be considered as opening a loophole, one called the detection loophole.

In fact this loophole is due to the use of less-than-ideal devices, here low-efficiency detectors. The situation is very different depending on the particle used during the experiment. For optical photons the efficiency of the detector is very low (10–30%). It is larger for gamma rays. The detectors for massive particles are generally more efficient.

In 2001, Rowe[25] et al. claimed that their experiment was the first one to close the detection loophole, and many authors[26] came to agree with them in the following years. Rowe et al. measured the quantum correlation between two Be ions with nearly perfect detection efficiency. But note that while the efficiency was 80% (large compared to photon detector efficiency), the locality loophole remained largely open in this experiment since the ions were separated by only 3 μm!

Since 2001, more experiments with massive particles have been performed. Recently, a violation of Bell's inequalities between a quantum state of two remote Yb^+ ions separated by a distance of about 1 m was shown.[27] However, a much

---

21 Aspect (1999).

22 Tapster et al. (1994).

23 Tittel et al. (1998).

24 Salart et al. (2008).

25 Rowe et al. (2001).

26 Grangier (2001).

27 Matsukevitch et al. (2008).

larger separation between the ions or a shorter detection time is necessary for a loophole-free test of a Bell inequality. For example, a realistic 50-μs detection time will require 15 km separation between the ions.

6.4.2.3 Some General Considerations Concerning the Closure of the Loopholes

When a significant improvement is reached towards the closure of a given loophole, new problems (very few of which have been taken into consideration in the past) are pointed out by some authors, leading to a new course for technical improvements. This is typically the case when, after closing the locality loophole, people suddenly became aware of the detection loophole, or, more recently, when people suddenly began taking gravity into account. When one loophole seems to be closed in a given experiment, the second one remains open. For technological reasons, the simultaneous closure of both loopholes seems difficult.

6.4.3 Another Conclusion at This Point

The robustness of an experimental result (here the violation of Bell's inequalities) inevitably raises the question of the solidity of the individual experimental derivations. In the case of the test of Bell's inequalities, the discussion of the solidity of the individual derivations includes a discussion on the ability of the real experimental test to reproduce the ideal "gedanken experiment". In fact, up to now, the real experiments have all been too far from the ideal one; and this leads to the loopholes described in the previous sub-sections.

The problem is that the experiments are either based on additional assumptions compared to the ideal experiments or else they do not fulfil all the requirements of Bell's theorem. Because of these loopholes, the various individual experiments, even if accurate and reproducible, can't be considered as correct (valid) experimental tests *of Bell's theorem*. Thus, some scientists think that the so-called locality and detection loopholes constitute a failure in the robustness of the conclusion that the Bell's inequality are indeed violated, and hence a failure in the robustness of the conclusion that SQM is indeed corroborated and LHVT refuted by the available set of experiments.

We are in fact presented with a large number of (partially) independent experiments with convergent outcomes and *with different assumptions (related to different loopholes)*. The important point here is the problem of the assumptions. Levins[28] discusses the robustness of a biological assertion that is validated by a large number of independent models that give invariant results. According to him, the fact that the models are based on different but more-or-less correct, precise but imperfect assumptions is not incompatible with the robustness of the biological assertion.

[28] Levins (1966).

But, as we will see in the next section, as far as imperfections are recognized in the assumptions involved in the different experimental derivations, the robustness of the experimental result hangs on an evaluation of the importance of such imperfections, and this evaluation is not always uniform among practitioners. In the case of the experimental tests of Bell's theorem, it is not. We can reconstruct the position of some scientists as a judgment of failure in robustness because of the imperfection in the assumptions used in the experiments.

6.5 Some Philosophical Questions Related to the Robustness of the Result of Tests of Bell's Inequalities

6.5.1 Are Some of These Experiments More Crucial Than the Others?

Considering Bell's inequalities tests, we are presented with several groups of experiments, every group using a particular experimental set-up or family of experimental set-ups (Fig. 6.2): the atomic cascade group, or PDC group or the massive particles group. In each group several experiments are performed successively, each new one being in general an improvement of the preceding ones. What is noticeable is that the successive authors claim that they have done a better experiment than the previous one.

Considering each loophole separately, it seems that, among the large number of experiments performed, there are experiments that are considered more crucial than the others as a step towards the closure of a given loophole. These experiments are considered as the "best ones" – until another one shows a significant improvement towards the closure. The arrows linking these experiments to their outcomes are the bolder arrows in Fig. 6.2 (the bigger the arrow, the most decisive is the experiment according to most physicists today).

Concerning the locality loophole, Aspect's experiment in 1982 and Weihs' in 1998 have successively been considered as decisive. It is noteworthy that today in the French physics community, the Aspect experiment is still presented as "*the* crucial experiment",[29] the following work amounting to just "small" improvements. For many physicists, this experiment has solved the locality loophole, and the detection loophole doesn't need consideration. However, Aspect himself claimed in 1999[30] that the locality loophole was in fact closed only in 1998 by Weihs and co-workers. His argument was that this experiment *exactly* follows the scheme of an ideal timing experiment.[31]

[29] *Les dossiers de la Recherche* (2007).

[30] Aspect (1999).

[31] Aspect (2004).

Concerning the detection loophole, the Rowe experiment done in 2001 has been considered decisive by a large part of the community of physicists.

6.5.2 Is the Detection Loophole as Relevant as the Locality One? Do We Need to Close It?

What is the position of scientists concerning the detection loophole? In fact, depending on the physicist or on the philosopher, and depending on the time, little or great importance is devoted to the loopholes.

The low efficiency of optical-photon detectors has always been widely recognized, but it has been generally considered a minor practical problem. One of the reasons is that Bell himself considered it acceptable to make a fair sampling assumption, that is, to extrapolate the results actually obtained in the experiment with low efficiency detectors to detectors 100% efficient. The fair sampling assumption was justified by Bell in 1981 with the following frequently quoted sentence: "It is hard for me to believe that SQM works so nicely for inefficient practical set-ups and is yet going to fail badly when sufficient refinements are made". Naturally, this official position of a famous physicist had a great influence on the judgment of the scientific community.

Let's quote Aspect, as an illustration of a physicist's position (probably influenced by Bell) on the question of the detection loophole:[32]

> This experiment [the Weihs' experiment] is remarkably close to the ideal 'gedanken' experiment, used to discuss the implications of Bell's theorem. Note that there remains another loophole, due to the limited efficiency of the detectors, but this can be closed by a technological advance that seems plausible in the foreseeable future, and so does not correspond to a radical change in the scheme of the experiment. Although such an experiment is highly desirable, we can assume for the sake of argument that the present results will remain unchanged with high-efficiency detectors.

Another reason why the detection loophole is considered a minor loophole by physicists is that it is a usual loophole, a usual problem that exists in many physics experiments. Considering the detection loophole to be a serious problem in the case of Bell's inequalities tests would have an important consequence: a loss of confidence in the results of a lot of experiments.

However, the detection loophole presents some specific issues when considering the experimental tests of Bell's inequalities, and some authors have considered this problem serious. Among them, Santos[33] argues that the core of hidden variable theories is that systems that are identical according to SQM are actually not identical. Each of them has its own hidden variable. Obviously, particles with different hidden variables will behave differently, and this includes the whole process of crossing

[32] Aspect (1999).

[33] Santos (1992).

the polariser and being detected. Santos concludes that, although a linear extrapolation of the measured polarization correlation to higher efficiency would violate the Bell's inequalities, this fact does not imply that a violation will be produced if the experiments are actually performed with more efficient detectors.

In the same way, Clauser and Horne, Gisin and Percival[34] underline the fact that we have to make an additional assumption in order to obtain the probabilities that appear in the inequalities: this assumption is that *the detector efficiency is independent of the local hidden variables*. In any case, as underlined in Section 6.4, some experimentalists have considered the detection loophole serious enough to undertake specific experiments with massive particles in order to close it.

6.5.3 Is an Ultimate Experiment Closing Simultaneously Both Loopholes Desirable? Is It Necessary?

Many people obviously think a loophole-free test is still desirable. To be convinced of this assertion, one has to remark that over the years several experimental schemes for a loophole-free test of Bell's inequality have been proposed[35] in the literature. In fact, a proposal for a loophole-free test with electron spins of donors was made as recently as March 2008.[36] And recently a new experiment with a Yb ion was performed with the aim of closing both loopholes.[37]

Among physicists and philosophers, three different positions are found with respect to the necessity of a loophole-free experiment: an intermediate position (i) according to which an ultimate experiment is desirable but not necessary, a strong position (ii) from people open to the LHVT possibility and a weak position (iii) from people convinced that the LHVT is already refuted and that new experiments will inevitably lead to the same conclusion as the already performed ones.

(i) Some scientists think that an ultimate experiment is desirable although not necessary. These scientists are very confident that such an experiment would favour SQM. This intermediate position is well illustrated by the writings of Genovese:[38]

> A conclusive experiment falsifying in an absolutely uncontroversial way local realism is still missing... In the 40 years various experiments have addressed this problem: strong indications favouring SQM have been obtained, but no conclusive experiment has been yet performed, mainly due to the low efficiency that demands for additional assumptions. Nevertheless, relevant progresses have been made in the last ten years and in my opinion an ultimate experiment could be far in the future.

[34] Percival (2000b).

[35] Kwiat et al. (1994), Huelga et al. (1995), Fry et al. (1995), Garcia-Patròn et al. (2004).

[36] Hong et al. (2008).

[37] Matsukevitch et al. (2008).

[38] Genovese (2005).

Ph. Grangier,[39] a famous physicist involved in performing an experimental test of Bell's inequalities, also defends an intermediate position:

> Closing both loopholes in the same experiment remains a challenge for the future, and would lead to a full, logically consistent rejection of any local realistic hypothesis. Even so, the overall agreement with quantum mechanics seen in all the experimental tests of Bell's inequalities is already outstanding.
>
> Moreover, each time a parameter is changed that was considered to be crucial (for example, using time-varying measurements, or increasing the detection efficiency), the experiments show that these changes have no consequence: the results continue to agree with quantum-mechanical predictions. This appears rather compelling evidence to me that quantum mechanics is right, and cannot be reconciled with classical physics.

As we have seen above, this position is also shared by Aspect. However, this intermediate position is not shared by everyone.

(ii) On one side, some authors claim that the lack of a conclusive experiment after 40 years could indicate the practical impossibility of falsifying LHVT and that in any case, we don't have enough elements to endorse a neat and definitive conclusion at this stage. For example, according to Santos,[40] a researcher in theoretical physics, the fact that loopholes appear in every experiment is an argument in favour of adopting a position open to local realism (instead of the widespread position that the issue is already settled):

> A serious attention to the loopholes in the empirical tests of the Bell's inequalities, rather than their uncritical dismissal, may improve our understanding of nature. In any case, the validity of local realism may be either refuted by a single loophole free experiment or increasingly confirmed by the passage of time without such an experiment. The extreme difficulty to make a loophole free test, proved by its unsuccessful effort of 40 years does not support the common wisdom that the question of local realism is settled. On the contrary the conclusion is that further research is needed.

So advocates of both the intermediate position and the strong position think that a loophole-free experiment is desirable. Both extrapolate the outcome of this possible future experiment but in different directions: Grangier is confident the outcome will again support SQM (so even if desirable, the ultimate experiment is not necessary), whereas Santos imagines that the outcome could favour LHVT and thus that such an experiment is necessary.

(iii) On another side, some authors deem that the large amount of experimental data disfavouring LHVT is already largely sufficient to exclude LHVT. In particular, they don't think that the detection loophole merits additional attention. They don't think an ultimate experiment is necessary or even desirable. This is the position of John Bell, among others.

[39] Grangier (2001).

[40] Santos (2005), p. 562.

6.5.4 The Ultimate Experiment and Robustness

At the end of the day, is the result of the multiple experiments performed so far in order to test Bell's inequalities and SQM against LHVT robust, and is it robust enough to dismiss the need of further improved experiments? Actually, as shown in the previous section, the answer to this question varies according to the scientists. Advocates of the intermediate position (i) and of the weak position (iii) think that the result of the experimental tests of Bell's inequality is robust in Wimsatt's sense. However, the degree of robustness they attribute to the experimental result is not the same. The scientists of the first group think that the robustness of the result might be increased and that this increase (and hence a loophole-free experiment) is desirable. Whereas scientists of the second group think that the result is solid enough that an additional experiment is useless. As for advocates of the strong position (ii), they don't agree with the idea that the result of the multiple experimental derivations is robust. To their eyes, the result is still fragile. They claim that only a loophole-free experiment, if its outcome were indeed convergent with the outcomes of the already performed experiments, would lead to a robust result.

6.6 Conclusion

From the early 1970s to the present day a large number of experimental tests of Bell's inequalities have been performed. The outcomes – except one – favour SQM and are inconsistent with LHVT.

At a first glance, the result of the experimental tests of Bell's inequalities is robust if we follow the initial statement[41] of Wimsatt: a result is robust if it is invariant under a great number of independent derivations. However, in the present case, we could argue that we are in presence of an 'apparent robustness'. This is because real experiments differ from the ideal experiment used to derive the Bell's theorem in several respects. To recall the two main kinds of problems:

1) In all the experiments an additional assumption is used due to the fact that a part of the experimental set-up is not 100% efficient. This leads to the detection loophole.
2) The experiments do not fulfil one of the requirements of Bell's theorem, for example the locality condition. This leads to the locality loophole.

Consequently, strictly speaking, one cannot conclude today that the experimental tests have ruled out the LHVT. And as we have seen, some scientists consider that the result of the multiple experiments performed so far is not robust, or not robust enough: they call for an ultimate experiment that will increase the robustness of the result.

[41] Wimsatt (1981).

Our conclusion is that a failure of the robustness of the result of the Bell's inequalities experimental tests is induced by the loopholes. We state that, in order to conclude that an experimental result is robust (here conceived as a test of a given theoretical question), one has also to consider the solidity of the various independent individual derivations carefully. The solidity of the individual derivations includes their accuracy and reproducibility, but also the ability of the real experiment to be a good instantiation of the "gedanken experiment", without additional or missing assumptions with respect to the ideal experiment.

We would like to emphasize that the above failure of robustness is completely different from the prototypical failures usually put forward in relation to Wimsatt's robustness scheme. The prototypical case corresponds to the three following possibilities: a discrepancy between (or a lack of convergence of) the multiple outcomes of the multiple derivations; or too few derivations; or a lack of independence between the derivations. The type of failure we have stressed in this chapter occurs at the level of the individual derivations, because the available derivations, although indeed multiple, convergent and in several important respects independent, are perceived as not solid enough.

An additional conclusion of this chapter concerns the lack of convergence of the scientists' own stated positions. From the analysis of the numerous comments of both experimental and theoretical physicists and of philosophers, one can conclude that they do not agree on several questions: on the importance of the detection loophole; on the decisive character of experiments and hence on the necessity to perform an ultimate experiment. Thus, depending on the authors, the results of the experimental tests of Bell's inequalities are today more or less robust.

A similar discrepancy of the judgments inside the scientific community about the validity of an experimental result has already been emphasized, for example by Harry Collins.[42]

Finally, I would like to recall that we have been concerned here with the battle of SQM against LHVT, that is *local* hidden variable theories. We need to keep in mind that non-local hidden variable theories still remain a possible alternative to SQM. For example, Bohmian mechanics and stochastic approaches are not invalidated by the above experimental tests, since they are non-local theories.

Acknowledgement The author would like to thank Soler and Trizio for helpful comments.

References

Aspect, A. 1999. "Bell's Inequality Test: More Ideal than Ever." *Nature* 398:189–90.

Aspect, A. 2004. "Bell's Theorem: The Naive View of an Experimentalist." ArX iv:quant-ph/0402001.

Aspect, A., Ph. Grangier, and G. Roger. 1982. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedanken Experiment: A New Violation of Bell's Inequalities." *Physical Review Letters* 49:91–4.

[42] Collins (1985, 2002).

Bell, J.S. 1964. "On the Einstein Podolsky Rosen Paradox." *Physics* 1(3):195–200.

Bell, J.S. 1987. *Speakable and Unspeakable in Quantum Mechanics*, 14–21. Cambridge: Cambridge University Press.

Berche, B., C. Chatelain, C. Dufour, T. Gourieux, and D. Karevski. 2006. "Historical and Interpretative Aspects of Quantum Mechanics: A Physicists' Naive Approach." *Condensed Matter Physics*. 2(46):319.

Collins, H. 1985, 2002. *Changing Order: Replication and Induction in Scientific Practices*. Chicago: University of Chicago Press.

Einstein, A., B. Podolsky, and N. Rosen. 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47:777–80.

Freedman, S.J., and J.F. Clauser. 1972. "Experimental Test of Local Hidden-Variable Theories." *Physical Review Letters* 28:938.

Fry, E.S., et al. 1995. "Proposal for a Loophole-Free Test of the Bell Inequalities." *Physical Review A* 52:4381–95.

Garcia-Patròn, R., et al. 2004. "Proposal for a Loophole-Free Bell Test Using Homodyne Detection." *Physical Review Letters* 93(13):130409.

Genovese, M. 2005. "Research on Hidden Variable Theories: A Review of Recent Progresses." *Physics Reports* 413:319.

Grangier, Ph. 2001. "Quantum Physics: Count Them All." *Nature* 409:774.

Hong, F.Y., et al. 2008. "Proposal for a Loophole-Free Test of Nonlocal Realism with Electron Spins of Donors.". *Physical Review A* 77:052113.

Huelga, S.F., et al. 1995. "Loophole-Free Test of the Bell Inequality." *Physical Review A* 51: 5008–11.

Kwiat, P.G., et al. 1994. "Proposal for a Loophole-Free Bell Inequality Experiment." *Physical Review A* 49:3209–20.

Lakatos, I. 1980. "The Methodology of Scientific Research Programmes." In *Philosophical Papers*, vol. 1, edited by J. Worrall and G. Currie. New York: Cambridge University Press.

Les dossiers de la Recherche No. 29, 2007. "Le monde quantique".

Levins, R. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54:421.

Matsukevich, D.N., P. Maunz, D.L. Moehring, S. Olmschenk, and C. Monroe. 2008. "Bell Inequality Violation With Two Remote Atomic Qubits." *Physical Review Letters* 100:150404.

Percival, I.C. 2000a. Speakable and Unspeakable After John Bell. A talk given at the International Erwin Schrödinger Institute, Vienna (ESI) at the November 2000 Conference in commemoration of John Bell 2000 Dec 05.

Percival, I.C. 2000b. "Why Do Bell Experiments?" *Nature* (submitted). arXiv:0008097v1.

Rowe, M.A., et al. 2001. "Experimental Violation of a Bell's Inequality with Efficient Detection." *Nature* 409:791–4.

Salart, D., et al. 2008. "Space-like Separation in a Bell Test Assuming Gravitationally Induced Collapses." *Physical Review Letters* 100:220404.

Santos, E. 1992. "Critical Analysis of the Empirical Tests of Local Hidden-Variable Theories." *Physical Review A* 46:3646.

Santos, E. 2005. "Bell's Theorem and the Experiments: Increasing Empirical Support for Local Realism?" *Studies in History and Philosophy of Modern Physics* 36:544–65.

Tapster, P.R., J.G. Rarity, and P.C.M. Owens. 1994. "Violation of Bell's Inequality over 4 km of Optical Fiber." *Physical Review Letters* 73:1923.

Tittel, W., J. Brendel, H. Zbinden, and N. Gisin. 1998. "Violation of Bell Inequalities by Photons More than 10 km Apart." *Physical Review Letters* 81:3563.

Weihs, G., T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger. 1998. "Violation of Bell's Inequality Under Strict Einstein Locality Conditions." *Physical Review Letters* 81:5039.

Wimsatt, W.C. 1981. "Robustness, Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins, 123–62. San Francisco, CA: Jossey-Bass Publishers.

Chapter 7
Scientific Images and Robustness

Catherine Allamel-Raffin and Jean-Luc Gangloff

As Léna Soler[1] has emphasized:

> (...) the term 'robustness' (...) is, today, very often employed within philosophy of science in an intuitive, nontechnical and flexible sense that, globally, acts as a synonym of 'reliable', 'stable', 'effective', 'well established', 'credible', 'trustworthy', or even 'true'.

But in parallel, William C. Wimsatt has developed a specific sense (Wimsatt 1981, 2007), which, while preserving the common association with the ideas of reliability and effectiveness, is more precise and more technical, and refers to the idea of the invariance of a result under multiple independent derivations. In this paper, we argue that "robustness analysis" (in Wimsatt's sense) is nothing less than the guiding principle of the argumentative structure of many papers published in natural sciences. We base our analysis on the methodology of ethnographic studies. Our aim is to take into account the actual practices which occur in laboratories (Allamel-Raffin 2004, 2005).[2] Our approach is mainly descriptive, although it does not exclude a normative perspective; for we conceive norms to be elaborated in the research process. In other words, we believe that norms are historically set up. Besides, we think that problems raised by philosophers are also faced by scientists. For example, "the experimenter's regress" or "the theoretical underdetermination by the data", as Kitcher (2001) observed are not only issues identified by philosophers but also by scientists themselves. Our argument is based on the examination of a 2001 astrophysical paper: "The Milky Way in molecular clouds: a new complete CO survey".[3] One of our purposes is to link the "robustness analysis" with the use of images in scientific papers. We shall see how images can never be reduced to mere illustrations but are an important component of the argumentation, and thereby

[1] See Chapter 1, p. 3.

[2] Our study relies on an ethnographic investigation conducted in an astrophysics lab: the Harvard-Smithsonian Center for Astrophysics (CfA), Cambridge, USA.

[3] Dame et al. (2001).

C. Allamel-Raffin (✉)
IRIST, University of Strasbourg, Strasbourg, France
e-mail: allamelraffin@unistra.fr

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_7,

of the robustness of the results. First, we shall focus on the various discourses on robustness in philosophy of science in order to show that Wimsatt's "robustness analysis" concept is better able to capture laboratory practices but also the structure of a scientific paper (Sections 7.1 and 7.2). In a second part, we study how generally a scientific paper is not a demonstration, strictly speaking, but rather a sequence of arguments (Section 7.3). For that purpose, we examine the argumentation of the aforementioned astrophysical paper and especially the role played by images in the argumentation (Sections 7.4, 7.5, and 7.6). Finally, insofar as the notions of independence and invariance are crucial to the concept of "robustness analysis", we shall comment on them relying on our case study on the Milky Way (Section 7.7).

7.1 Robustness Analysis of Philosophers

The concept of "robustness" has been explicitly thematized in philosophy of science for three decades. As with any concept, it's a "working" concept, and we can use it in different contexts. And as with most empirical concepts, this one has an "open texture", as argued by Waismann (1945). Empirical concepts are not defined as a set of characteristics established once and for all (as in the formal sciences). They change over time, and one can imagine that their new characteristics are influenced by new practices or new conceptual frames. This is probably the reason why "robustness" has generated different definitions. Several authors have proposed different versions of the concept of "robustness" and illustrated them by different examples. For example, in a case study of DNA sequencing, Culp (1995) shows how two completely different methods led to comparable results. Nederbragt (2003) presents a case study of the invasion of cells by microorganisms. He also shows that scientists have used two or three different methods or instruments in order to confirm their theory. In a historical study about the reliability of thermometers, Chang (2001, p. 283) explains that the use of several thermometers – based on different principles: air, carbonic acid, hydrogen – led the physicist Regnault to close the debate about temperature. Chang insists on showing that in this case there was no appeal to any theory to close the discussion. Ian Hacking, for his part (1981, pp. 144–145; 1983, pp. 324–332) takes the example of the dense bodies in red blood platelets. He shows that these entities can be detected by two different microscopes relying on different properties of light, namely, the transmission electron microscope and the fluorescent microscope. Hacking[4] terms this "the argument from coincidence".

[4] He also develops another concept: the "robust fit". In a sense, this concept is close to robustness analysis. What is a "robust fit"? We will mainly rely on the definition proposed by Hacking (1992). According to this view, the "robust fit" is the adjustment, in the laboratory, between three fundamental categories (ideas, things and marks) that enable scientists to obtain the reliability and the repeatability of the results. The concept of "robust fit" is slightly different from the robustness analysis, but what is in common, is the procedure consisting in *crossing several elements* (theories, instruments, know-how, etc.) to obtain robust scientific results.

Taking into account all these analyses, Nederbragt affirms that several names were given to the same procedures consisting in 'using different methods to confirm an hypothesis'. These denominations can be: "robustness analysis" (Wimsatt 1981), "triangulation" (Culp 1995; Wimsatt 1981), and "independence of the routes" (Hudson 1999). In our eyes, all these definitions have something in common: they fit Wimsatt's definition, which is:

> The family of criteria and procedures which I seek to describe in their various uses might be called *robustness analysis*. They involve the following procedures: 1/ To analyse a *variety* of *independent* derivation, identification, or measurements processes. 2/ To look and analyse things which are *invariant* over or *identical* in the conclusions or results of these processes. 3/ To determine the *scope* of the processes across which they are invariant and the *conditions* on which their invariance depends. 4/ To analyse and explain any relevant *failures* of *invariance*. I will call things which are invariant under this analysis "robust". (Wimsatt 1981, p. 126)

For us, this definition is by far the more precise and technical. It's the reason why we will rely our analysis on it.

7.2 Robustness Analysis in Scientific Laboratories

What was extremely striking during our ethnographic studies is how the robustness analysis is entrenched in the day-to-day activities of an astrophysical laboratory such as the Center for Astrophysics in Harvard. Robustness analysis is a fundamental methodological principle because the astrophysicists' enquiry has to deal with two dimensions of observability: directness and amount of interpretation (Kosso 1989). According to Peter Kosso, directness can be understood as a dimension of observability: there are more or fewer interactions from the source X observed conveying an information to the final human receptor. Directness, according to Kosso, is a measure of the physical closeness between the source and the final receptor. In contemporary astrophysics, the chain of interactions included in an observation is often long and complex and the observation consequently hugely indirect. Another dimension of observability is the amount of interpretation: how many distinct physical laws are needed to get from the source X to the final receptor. The amount of interpretation is a measure of epistemic closeness between this source and the final receptor. The number of such laws, in astrophysical observations, is often very high. In absolute terms, for one observation, the scientists should be able to make explicit what kind of physical interactions occurred and which physical laws are involved in the observation. They should also be able to make explicit how they identified and quantified the noise. What makes things a lot more difficult is the possible existence of artefacts (the sources of certain of these artefacts being unknown) and the existence of tacit knowledge due to the fact that this information is produced by human beings. To reduce these difficulties, one can think that the solution must be found in instrumentation, in the sense that instruments perform a lot of tasks automatically. But that would be too simple... The old ninenteenth century ideal of "mechanical objectivity" (Daston and Galison 1992, 2007) and consequently the hope that scientists' subjectivity could be eliminated, has vanished. We know that we can't avoid the presence of human subjectivity in the use of instruments. We know that the use of each instrument includes subjective choices and also skills, the canonical example being in this case photographical techniques. Idiosyncratic characteristics are always present and cannot be definitively eliminated. The only way to get out of this impasse is to recognize that objectivity comes in degrees (Culp 1995) and has to be conceived more as a *continuum* (Putnam 2003). For those who agree with this view, the goal is then to reduce idiosyncratic characteristics as much as possible.

In the end, we can recognize especially if we stay a long time in a lab,[5] that we have an interpretative flexibility of the data as Harry Collins (1981) describes in his papers, but:

a) we can notice that the scientists are perfectly aware of this interpretative flexibility of the data;
b) Therefore, the scientists try to reduce this interpretative flexibility. The robustness analysis seems for them a *practical necessity.* They aim above all to produce 'robust' results.

7.3 Argumentation in Scientific Papers

It follows from these ideas that numerous scientific papers do not "demonstrate", but propose an argumentation. The astrophysicists get with their data or images only what we can call "pieces of evidence" or "elements of proof". They do not start with true premises to get in the end, by following the rules of formal logic, absolutely true conclusions. Thereby, they are forced to propose in their papers the more convincing way to expose their pieces of evidence or elements of proof. What is argumentation? Briefly, it includes the following features[6]: any argument contains a claim, an assertion put forward publicly for general acceptance. This claim is supported by grounds or statements specifying particular facts about a situation. These facts are already accepted as true and can therefore be relied on to clarify and make good the previous claim (establish its truth or correctness). Sometimes, we need warrants, statements indicating how the grounds or facts on which we agree are connected to the claim. These connecting statements draw attention to the previously agreed general ways of arguing applied to the particular case and so are implicitly relied on

[5] This conclusion is drawn from our ethnographic studies. We spent three months in the Harvard-Smithsonian Center for Astrophysics (USA), six months in a nanoscience laboratory (France) and three months in a pharmacology laboratory (France). In each case, we observed numerous times this phenomenon of interpretative flexibility, especially when the scientists work on scientific images. For example: see Allamel-Raffin (2005).

[6] Our main references concerning the question of argumentation are the following books: Toulmin (1958) and Janik et al. (1984). We are perfectly aware that a lot of literature exists on the argumentation topic, but we have chosen to use the concepts of Toulmin & *al.*, because they are useful enough to help us to understand the general argumentation of a scientific paper.

as those whose trustworthiness is well established. In scientific papers, the critics can focus on the grounds, the data itself (arguing that it is false). But they can also ask questions about the use of a given instrument or a given technique employed to obtain this data. Sometimes, the warrant has to be justified, and that is the function of the backing, the generalization making explicit the body of experience relied on to establish the trustworthiness of the ways of arguing applied in any particular case. With this definition, we'll examine in our two next sections the astrophysical paper's argumentation.

7.4 A Case Study[7]: A New Complete CO Survey in the Milky Way

The paper's title is: "The Milky Way in molecular clouds: a new complete CO survey". The paper is a perfect illustration of the work done in a context of normal science activity. It's a radio astronomy[8] paper, one of the most cited papers in radio astronomy since it was published in 2001. The article includes two big maps (1.5 m long). These maps constitute the main result of the paper and represent the CO distribution in our galaxy (Fig. 7.1).

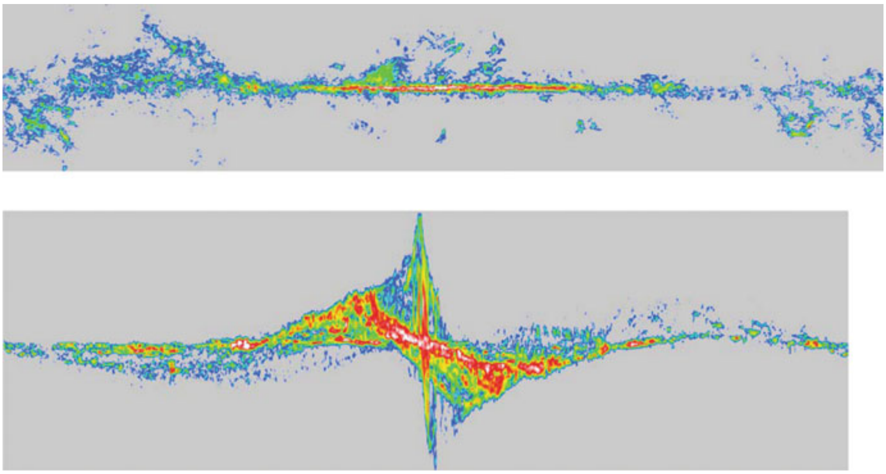

Fig. 7.1 Two big radio maps of the Milky Way (1a and 1b)
Source: Dame et al. (2001, CfA)

[7] Our analysis relies not only on the paper itself. It also includes some ethnographic material collected during our stay in the Center for Astrophysics at Harvard: interviews recorded with one of the paper's authors (Thomas Dame) and observation reports of the day-to-day activities in the lab.

[8] Radio astronomy is a subfield of astronomy that studies celestial objects in the radio frequency portion of the electromagnetic spectrum, that is to say the wavelength between 0.3 and 2500 mm.

Why is it interesting to study the presence of CO in the Milky Way? In fact, CO detection is a mean to study dust clouds. These dust clouds are very interesting for many reasons. One of these reasons is that they are the birthplace of new stars. Dust clouds are made of molecular gas and atomic gas. There are many molecular gases, H_2 being the most important. But H_2 is extremely difficult to detect from the earth. CO is also a molecular gas, and many studies show that it's mixed with H_2 in dust clouds. So CO is a tracer of H_2 and of the dust clouds in the interstellar medium. CO is relatively easy to detect (its wavelength is 2.6 mm). The purpose of the two maps in the paper is to be useful for many other studies such as the aforementioned studies of the birth of stars, studies of the source of the cosmic rays, studies on the structure of our galaxy, etc.

Two associated claims are defended in the paper: the CO maps of the Milky Way:

(1) are constituted of reliable data
(2) are complete (there is no lack of data concerning the Milky Way).

The scientists want to argue that they have good quality data (no noise, no artefact). In other words, they argue that the data in the maps corresponds to reliable data. In the case, the data is about CO clouds. Furthermore, they argue that their two maps are complete (a map with some missing pieces would be useless). Our first preliminary analysis will focus on the arguments presented to support those two claims. What are the arguments exposed in the subsections of the paper in order to defend what we call the 'reliability claim'?

In Sections 7.4.1 and 7.4.2, we shall briefly analyze the structure of the paper and show that it fits very well with the robustness analysis of W. C. Wimsatt.

7.4.1 The Reliability Claim

7.4.1.1 First Argument: Analysis of the Instrumentation and of the Data Processing in Order to Justify Their Reliability

The paper begins with a description of the two telescopes used, including a brief history, and a discussion of the data acquisition and reduction employed in the various surveys. The scientists detail carefully the calibration procedures. A central point is constituted by the synthesis of the data. The survey was constructed from 37 individual surveys. It was crucial for the astrophysicists to explain how they managed to synthesize all these surveys, and more specifically, how they managed to reduce the noise included in these different studies. The way they suppressed noise is especially explained here because they didn't use the usual procedure to reduce noise in radio astronomy. If we compare the selected paper with other papers in astrophysics, this section is more developed than the same section in a current astrophysics paper. Briefly, the telescopes are two small millimeter-wave telescopes (1.2 m telescopes), one at the CfA (Cambridge Massachussetts), the other at the Cerro Tololo InterAmerican Observatory in Chile.

About the data acquisition, a first survey was realized between 1979 and 1987. This survey was published in 1986. The data acquisition[9] went on until 2000, always using the same two telescopes. The results (the maps) of the first survey were integrated to the new survey. The final result here is the CO maps. Theses maps are in fact the combination of 37 different small surveys. How did the scientists get their data ? It was a meticulous activity lasting more than 20 years. One of the researchers told us what they actually did in order to obtain their data. We quote him:

> So, we look at one spot in the sky at a time in order to build up these images. Each observation, because the signal is very weak, typically takes 2 or 3 minutes. So, that is why it takes such a long time.

The result of these observations is data cubes, in other words files in FITS format. The data files tell you all you need to know: when the data was taken, what part of the sky, what range in velocity, what frequency, etc. If the data stays in that format, it would be unexploitable. This raw data has to be converted into images.

Once the data collected with the FITS format,[10] the scientists have to manage the noise. Because they wanted to put all 37 surveys into one map, the sensitivity should have been limited by the worst survey, and in particular by the noise of this worst survey. Generally, in radio astronomy, astrophysicists show a little bit of noise on the map because human eyes are good at picking out real things from noise. Sometimes, the most interesting features are almost in the noise. One way to deal with the noise and to reduce it is to smooth the maps. To smooth a map means to take, for each pixel, a weighted average of the pixels around. The same thing must be done for each pixel. As a consequence, the researcher gets a fuzzier image, but it is more sensitive to anything that is extended. On the other hand, he can lose some very strong sources. That is a problem because what the scientists want in our case is to obtain a map – including very weak and very strong CO radio sources. So they couldn't use the traditional method of reducing noise. In order to resolve this problem, the astrophysicists employed a technique called "moment analysis". They took the whole data cube, and they first smoothed it quite heavily. What they did then was to degrade the resolution. This gave them a greater sensitivity for anything that is extended. They then used this smoothed data cube to reduce the noise: in any place in the data cube where there is no emission or where they think there is no significant emission, they blanked it in the original data cube. Basically, they used the smoothed data cube to tell them the regions of the data cube where there might be real emission

[9] In this section of the article, a special part is dedicated to the calibration procedures. In order to be sure that their calibration was correct, the scientists chose two different methods of calibration. These calibration choices are justified by referring to other studies in radio astronomy. These calibration procedures have nothing special in regard of the calibration procedure used in other radio astronomy studies of the same kind.

[10] FITS stands for 'Flexible Image Transport System' and is the standard astronomical data format endorsed by both NASA and the IAU. FITS is much more than an image format (such as JPG or GIF) and is primarily designed to store scientific data sets consisting of multi-dimensional arrays (1-D spectra, 2-D images or 3-D data cubes) and 2-dimensional tables containing rows and columns of data. http://heasarc.gsfc.nasa.gov/docs/heasarc/fits.html.

or significant emission. And where there is no emission, they just blanked out every pixel in the original data cube. What they got was a moment cube, which had the original resolution as they hadn't degraded the resolution, but the noise was reduced. After that, they could integrate all the way through for all velocities without picking up much noise. Thus they obtained a map where the noise is reduced and where the sensitivity is non-uniform. All this required an enormous amount of work! After that the radio astrophysicists checked many of the weaker features of the integrated maps to assure that they corresponded to identifiable spectral lines in the corresponding spectra and not to base-line fluctuations or statistical noise.

After these calibration procedures and data processing, why did the scientists persist in their activities? At this point they had thoroughly detailed the conditions of the data collected and the way they processed it. Isn't that enough to convince the readers of the paper? The claim that the points on the maps correspond to reliable data is supported by the description they made of the instrumentation, data acquisition and reduction. But that's not enough.

The warrant isn't strong enough: the warrant, here, consists in asserting that the instrumentation and the data processing are reliable. But only the quality of the data could certify these points. So, we are in the experimenter's regress... (Collins 1992; Collins and Pinch 1993). The data is good if the instrumentation and the data processing are correct, but to certify that the instrumentation and the data processing were correctly done, you have to look at the data!

In fact, as Peter Winch says, the scientific investigator is involved in two sets of relations: first, with the phenomena he investigates and second with his fellow-scientists and the rules of his scientific community (Winch 2007, p. 84). The authors of this paper know perfectly well that even if they are confident in their result because the instrumentation and the data processing are correct, it will be not enough to convince their fellows. They know that because they have internalized that culture through training in their field. As Gingras and Godin (2002) notice, "this is why scientists can anticipate criticism" and in our case, they know perfectly that good use of instrumentation and correct data processing is far from being a sufficient argument to convince their peers. It is why the argumentation must go on: four arguments, all related to the robustness analysis, are presented.

7.4.1.2 Second Argument for the Reliability Claim

In the same region of the Milky Way, the astrophysicists compare their data with other data obtained with a much bigger radiotelescope, with a ten times better resolution. The name of this telescope is the FCRAO.[11] This study was carried out by

[11] FCRAO stands for Five College Radio Astronomy Observatory. It was founded in 1969 by the University of Massachusetts, together with Amherst College, Hampshire College, Mount Holyoke College and Smith College.

Heyer et al. in 1998. The data is concordant. Here we have an independent[12] survey, but the same type of instrumentation with better resolution.

7.4.1.3 Third Argument for the Reliability Claim

This argument is a comparison of the data obtained by the astrophysicists on well known celestial regions as Orion, Centaurus, etc. with data obtained on the same regions by other teams of scientists using either radiotelescopes or other telescopes like optical, X-rays, gamma rays telescopes. Here we have several independent surveys[13] and various instruments characterized by independent physical principles.

7.4.1.4 Fourth Argument for the Reliability Claim

Finally, they compare their data pertaining to the center of the galaxy with data obtained with an optical telescope. Indeed, the CO is one of the gases present in the dust clouds. Other studies showed that dust blocks the distant starlight emitted in the optical wavelength. If they took an image of the center of the Milky Way in optical wavelength, they should observe obscure areas corresponding to the dust clouds made partially of CO gas. In this case we have an independent survey made by various instruments characterized by independent physical principles.

7.4.2 The Completeness Claim

The scientists aim to present a 'complete' CO survey of our galaxy, that is to say, they ignore only an infinitesimal part of the CO existing in our galaxy. To support that claim, they appeal to other studies of the galaxy:

- the first is a radio survey. This study was a survey of H I in the galaxy. H I is an atomic gas very easy to detect in radio astronomy (21 cm of wavelength)
- the second one is an infrared survey using the IRAS[14] telescope, looking at the dust clouds

Why should they use these surveys? Dust clouds are made of atomic gas and molecular gas. If you know the dust distribution, you know the gas distribution. CO is well represented in molecular clouds. If you have the distribution of dust clouds[15] in our galaxy and you subtract from that the atomic gas,[16] you will get a map of the molecular gas that you can consider as a predictive map of the CO in the galaxy. The

[12] 'Independent' means here that the survey was made by a different team.

[13] 'Independent' means here, once again, that the survey was made by different teams. We shall elaborate on the notion of independence in Section 7.7.1.

[14] IRAS stands for Infrared Astronomical Satellite. It's the first ever space-based observatory to perform a survey of the entire sky at infrared wavelengths.

[15] Astrophysicists get the dust clouds data using the IRAS survey.

[16] Astrophysicists get the atomic gas data thanks to the radio survey on HI.

scientists compare their data to those. So in this case we have independent surveys, instrument with the same physical principle/instrument with independent physical principle.

If we examine the five arguments presented in that paper, it appears that four of them are clearly a matter of robustness analysis. Each argument corresponds to a new derivation and each derivation gives new information about the phenomenon under study. Logically, on the basis of their own argumentation, the scientists conclude that results are robust, that is to say the maps represent only "reliable data", and the maps are complete. This first quick analysis has been done to illustrate, through this particular example that seems to us representative of many others, that robustness analysis is indeed a widely used, effective guiding principle of argumentation in scientific papers. Scientific papers are often a sequence of arguments organized according to a robustness scheme.

7.5 The Role of the Images in Scientific Papers' Argumentation

The images in scientific papers have traditionally been considered as mere illustrations. It is the propositional content, in contrast, which has been considered as essential and self-sufficient. Philosophers in particular have often underestimated the place of the images in scientific activities and results, victims of their "language-using ethnocentrism" as William C. Wimsatt has called it in his paper "Taming the Dimensions – Visualisations in Science" (1990). We believe that we must reconsider the role of images and their epistemic value in scientific papers. We must rethink their function in the composition and presentation of the robustness of results. In fact, we shall see that they play "a central role in the structure and the organisation of the scientific text. They are in fact the core of the scientific text" (Jacobi 1985). By "image", we take into account all what is non-textual in a scientific publication. That includes pictures, maps, graphs, histograms and so on.

Can an image constitute an argument? An image is not an argument if we take into account only its content and its internal structure. To be an argument, an image needs a textual support. The text actualizes some of the predication's virtualities contained in the image. The scientist who argues establishes some constraints that are guidelines for the final interpretation made by the reader. If the images structure the argumentation of a scientific paper, the text is complementary in two senses:

- The text reviews all the processes used to produce the image (instruments, data acquisition and data reduction). This is the "relay function" as defined by Roland Barthes (1964).
- The text limits the sense/meaning of an image in the caption. This is the "anchorage function" (Barthes 1964).

Let us return now to the analysis of our astrophysics paper and focus on the role of the images in the argumentation.

7.6 The Role of Images in the Milky Way Paper

At this point of our paper, we wish to focus more precisely on the details of the argumentation. The Milky Way paper has 10 pages. We find 20 figures which are either maps or curves. Basically, when we read the paper, what do we notice? It is made of text and images. We can ask now, what are the exact functions of the images in the argumentation we've briefly presented in Section 7.4.

7.6.1 The Role of the Images in the Reliability Claim

7.6.1.1 First Argument: The Role of Images in the Analysis of the Instrumentation and Data Processing in Order to Justify Their Reliability

There are no images in this section except one, showing the spatial position in the sky of the 37 studies. This is purely informative. There are no images of instrumentation. If the telescopes present anything out of the ordinary, they would probably have shown it on a photograph or a scheme. But this is not the case here, the telescopes used are very common in radio astronomy. For the data processing, there is no need for images because these procedures are common in astrophysics. In this section, the only real problematic point is the processing of the noise. The scientists dedicate the next section to this point.

7.6.1.2 Second Argument: The Role of Images in Comparing FCRAO Data with the CfA Data

As we said, the general claim to support here is the reliability claim. However, as soon as the scientists adopt one more derivation – here the FCRAO data – this general claim entails a more specific claim. The comparison between FCRAO data and CfA data aims to show that the particular noise processing done by the CfA scientists hasn't distorted the information obtained from the celestial sources. But they immediately face a problem: how can they compare the huge quantities of data acquired with the two telescopes? The challenge here is to find a relevant way to be able to compare them. In order to convince their reader, the scientists decided to associate three images.

These three images with the associated text are the argument ("grounds" in Toulmin's words). The caption stabilizes the meaning, in accordance with the anchorage function defined by Barthes.

Figure 7.2 enables a comparison between:

– dissimilar images (a) and (b).
– similar images (b) and (c)

It's not an isolated image that builds the argument but the joint use of a number of images. Images are juxtaposed so that they can be seen together. This is an

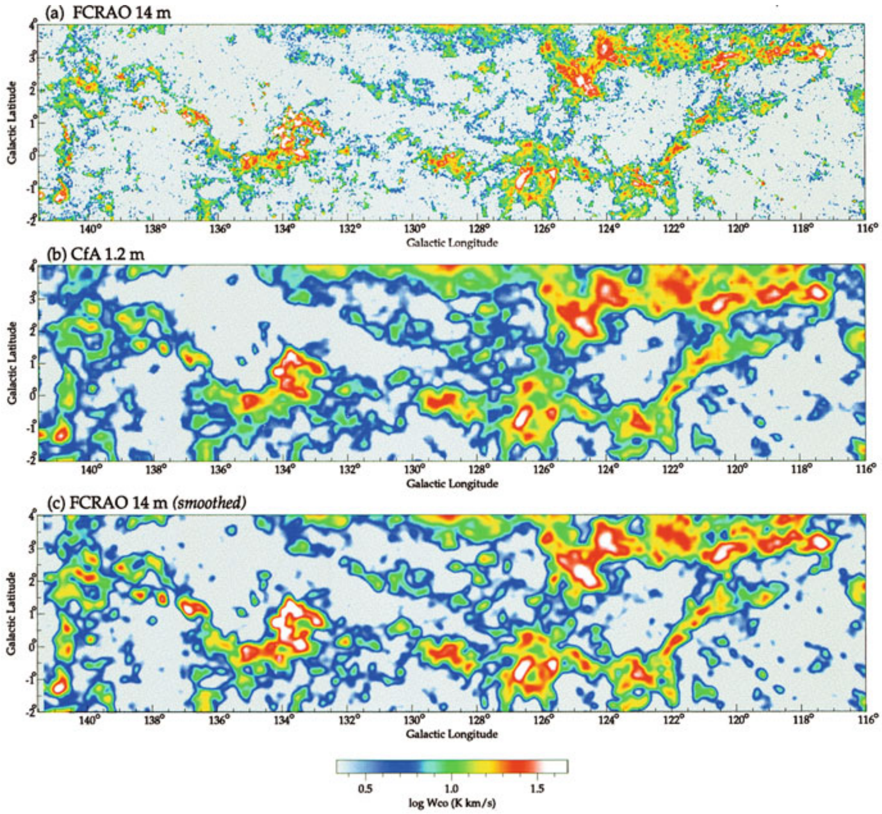

Fig. 7.2 (**a**) is the CO map of a specific celestial region. This map has been done by another team of researchers working with another radio telescope, the FCRAO, which has a ten time better resolution. (**b**) is the CO map of the same celestial region realized by the CfA radio telescope. (**c**) is the FCRAO map smoothed at the resolution of the CfA map. In accordance with its relay function, the main text describes the data processing from image (**a**) to image (**c**)
Source: Dame et al. (2001, CfA)

important point as emphasized by Tufte in his books (1990, 1997, 2007). Looking for differences and similarities requires this sort of comparative analysis. In our case, the argument consists in the visual similarities when one compares the images (b) and (c). Among the representational constraints, we have those relative to the best visualization. Using images enables:

– a better grasp of a huge number of data at the same time. It is necessary because of the overwhelming amount of data generated by the instruments.
– a better visualization of similarities or differences.

As one of the astrophysicists said about these images:

> This was my very first question about whether we analysed the data properly. (. . .) So, here is a comparison of the same region observed with two telescopes, independent telescopes.

> What I wanted to show is: everything that is on the map is real. (...) And then, you can see, amazingly well, these maps agreed. Even very small things, it's extraordinary. Better than I thought actually it would be. So, this was mainly to convince people that we've done everything right, because this is clearly independent data analysis. When you see that map, you believe it.

It would be unfeasible to compare directly the two sets of data produced by the two telescopes if they remain in numerical form. No one has the cognitive capacity to hold all items in the list in short-term memory, and then to do the calculations needed to extract conclusions about clouds' spatial localization and velocity. Taking some particular points on the map wouldn't be suitable either: in this case, one can always ask, what about the next point? Is it reliable? It would also be unfeasible to convert the relevant informational content of the map into a corresponding propositional content. Each map contains a huge number (potentially infinite) of predication's virtualities: there are the characteristics of each point and the links of each point with the others. Kitcher has stressed the same point about the Manhattan map (2001, p. 58):

> (...) the map is equivalent to a truly enormous number of claims about spatial relations: a picture is not worth a thousand words, but rather a staggering infinity of sentences. Further, although the map says many things that are incorrect, it also expresses an infinite number of true statements, for there are infinitely many truths of the form 'A is within Φ of being θ from due North of B', where A, B are places on the Manhattan shore and Φ, θ are angular measurements.

The fact that images can be potentially converted into a list of numerical data or into a propositional content doesn't mean that they are effectively converted in this kind of article. This is related to our limited cognitive abilities.

What is the strength of the astrophysicists' second argument?

If someone doesn't agree with the transition from these grounds (here the maps) to the claim (the reliability claim), he could ask for the warrant: the warrant here is that the different maps have been generated with two different radio telescopes in normal conditions of use. We find here a perfect example of the robustness' scheme in the sense of Wimsatt. This type of data processing is common in radio astronomy. If someone is not yet convinced by the warrant, he could ask for other general information to back up his trust in this particular warrant.

These telescopes and these procedures of data processing rely on well known physical theories. They are used in numerous studies without any problem. In the case of these three images, what is argued is: "Our data processing (and especially the processing concerning the noise) has not distorted the information you can see on the map".

7.6.1.3 Third Argument: The Role of Images in Comparing Optical Data with the CfA Data

Again, the general claim to support is the reliability claim. The comparison between the optical telescope and the CfA radio telescope aims to show that the

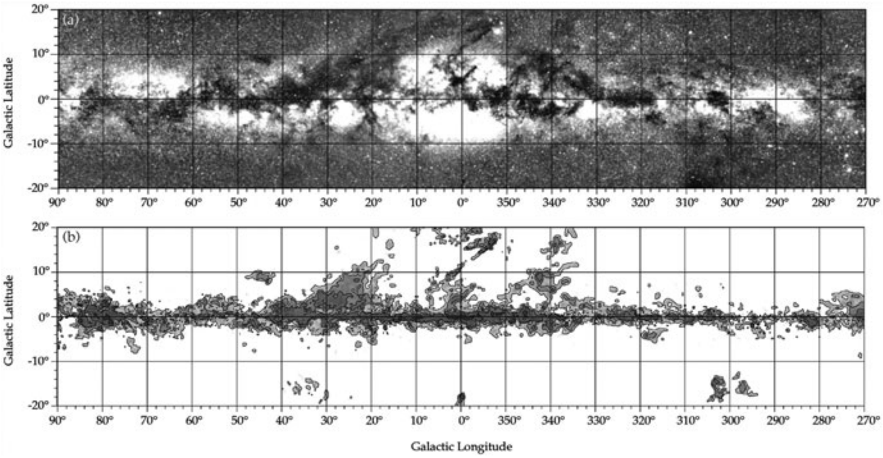

Fig. 7.3 (**a**) is a map created from 16 optical raw images of the galaxy. Like in radio astronomy, the procedure to put them together is called "mosaic". (**b**) is the same map as Fig. 1a, but it's zoomed on the center of the galaxy and put into grey scale
Source: Dame et al. (2001, CfA)

astrophysicists immediately meet the same problem: what can they do to compare the huge quantities of data acquired with the two telescopes in order to compare them effectively? They choose to create maps. In the Fig. 7.3, one can compare two reprocessed images.

The scientists zoomed into the center of the galaxy because it's the brightest region in which emissions are at their highest level in optical wavelengths. The aim is to show that the optical light is obscured by dust clouds, so it is the best region to do that. Usually in astrophysics, to compare two images with two different wavelengths, the researchers place an image over another image using, for example, contours. They didn't do that here. Why? Using contours was not a good way for representation, the visualization of the great similarities between these two images was not enhanced. Here is what one of the authors said:

> This was very challenging. It's just a mass of dark clouds. I tried white contours, coloured contours, nothing worked because anything you put on top of this, because it is a great correlation, you're getting CO emission where optical waves are. So I put a grid to help the eye, I couldn't do it another way.

The grid here is very useful to understand how good the similarities are between these two maps. To enhance the strong correlation between the two maps, the scientists use grey scales. The association of the two maps together with the caption is the argument.

Claim: Our radio map is similar to the optical one. This argues that our map shows only reliable features (as we know that CO is present in dust clouds).

Grounds: To assert this claim, we rely on the following observational data:

– one map in optical wavelength of the center of the galaxy
– our radio map of the same region

Warrant: the different maps have been created by means of two different telescopes in normal conditions of use. This type of data processing is current in radio astronomy and in optical astronomy.

Backing: These types of telescopes and these procedures of data processing rely on well known physical theories. They are used in numerous studies without any problems.

7.6.2 The Completeness Claim

The last argument we will examine is the one that supports the "completeness claim": "Our survey of carbone monoxyde in the Milky Way is complete". Astrophysicists used several independent surveys:

– IRAS infrared survey who detected the dust clouds in our galaxy
– A radio survey of HI gas in our galaxy. HI is an atomic gas.

Why should they use these surveys? As we already described in Section 7.4.2, dust clouds are made of atomic gas and molecular gas. From dust distribution, you know the gas distribution. A way to obtain a predictive CO map is to subtract the atomic gas data from the dust clouds data (cf. Section 7.4.2).

The argument here takes place in two figures: Figs. 7.4 and 7.5. The basic procedure here is again to show similarities in the maps.

To enhance the predicted CO, they turn it in white on the map for a better visualisation you can see the details much better.

The argument goes on in Fig. 7.5.

If you look at this figure, you can compare easily and understand that there is a strong correlation between the two maps. Again, in this figure, the scientists tried to enhance the similarities. They chose the same colour scale for the maps, they chose to reprocess their maps to extract a profile to be even more persuasive.

7.7 Some Remarks on Independence and Invariance

Independence and invariance are two notions that turn out to be very important in the concept of robustness analysis. Indeed, without a precise definition of these two notions, the concept of robustness analysis loses its significance. What could we say about these two notions if we take into account our present analysis?

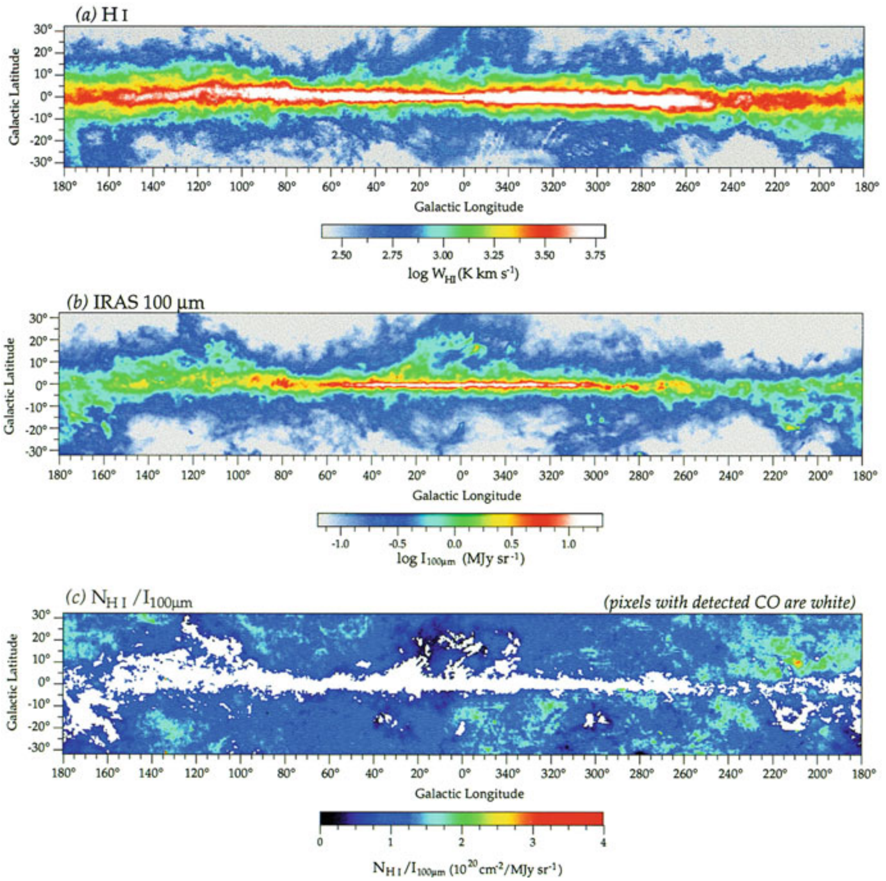

Fig. 7.4 (**a**) A radio map of the atomic gas (HI) in Milky Way done by another team with two other radio telescopes. (**b**) An infrared map of the dust clouds in the Milky Way done by another team with IRAS. (**c**) The IRAS map minus the atomic map (HI). It is in fact a CO predicted map *Source*: Dame et al. (2001, CfA)

7.7.1 Independence

Independence has to be understood in two different ways:

a) from a human perspective: From what the astrophysicist says in the interview, it is crucial that the data is produced by other teams of scientists. Why? Because scientists are aware of their own subjective choices. For example, in the noise treatment in our paper, it was crucial for the astrophysicists that the data processing didn't distort the data. So the fact that the radio data produced with

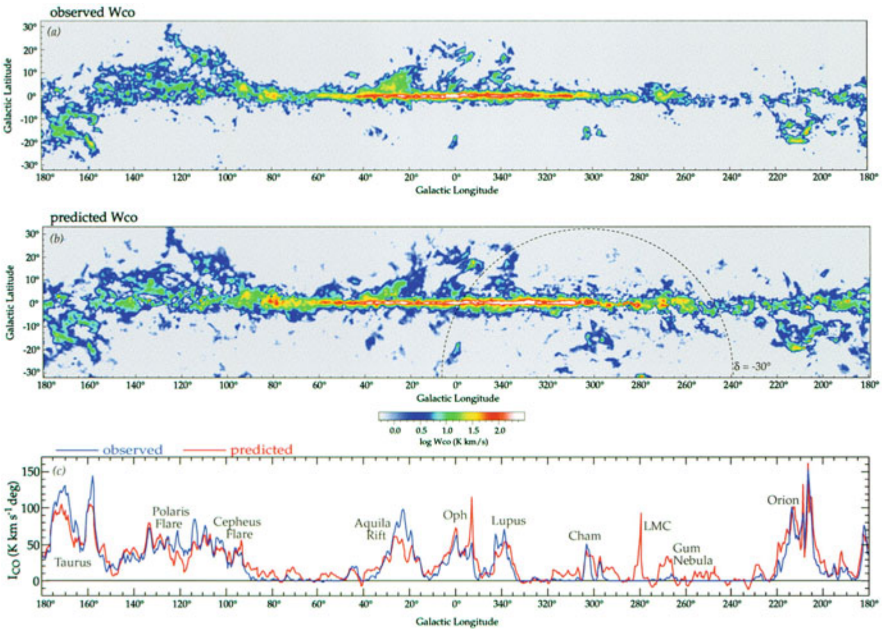

Fig. 7.5 (**a**) this is the CfA radio map. (**b**) the predicted CO map (the same that Fig. 4c but in another color scale). (**c**) is a more quantitative comparison. In *gray*: predicted CO and in *black*: observed CO. The *curves* are extracted from the two maps below
Source: Dame et al. (2001, CfA)

a better telescope (FCRAO) is similar allows them to be more confident in their own choices.

b) from an instrumental perspective: Astrophysicists use data produced by different types of telescopes (radio, infrared, optical telescopes). What does independence exactly mean when we talk about these different telescopes? In the case revealed in the Milky Way paper, telescopes record different wavelengths of the electromagnetic spectrum from radio wavelengths to gamma rays. One could say that the physical principles of these telescopes are not independent because they all allow the recording of electromagnetic radiations in the form of waves or particles. Each telescope allows the study of different properties of these molecular clouds: radio telescopes give information about the chemical properties of the molecular clouds; optical telescopes give information about the morphology of the molecular clouds, etc. In fact, the recording techniques and the technical problems are very different. These telescopes detect different properties of light (waves or particles). Each derivation (here different telescopes) enriches our understanding of the entity which is supposed to be the invariant under these multiple derivations.

7.7.2 Invariance

What is supposed to be invariant in our case? The localization and the velocity of the molecular clouds in the Milky Way, and more precisely, the localization and the velocity of CO which is located in these clouds. The spatial localization and the velocity are properties of an entity – molecular clouds – which are known by the detection of their properties. Different parts of the electromagnetic spectrum give us knowledge about different properties possessed by molecular clouds. So these clouds could be considered as entities which are clusters of properties. In fact, we could say that it is always properties and not objects that we primary observe – and not only in astrophysics. "Properties are primary, both metaphysically and epistemologically." (Humphreys 2006, p. 23). For Paul Humphreys, the ontological priority of properties suggests that the appropriate kind of realism to adopt here is property cluster realism. That kind of conceptualization seems to be a correct way to understand the way the scientists work. "The discovery of scientific entities thus involves a process rather like geographical discovery".

> (. . .) first an indistinct property is seen on the horizon; then some of the properties composing the coastline are discovered (. . .) then more and more details are filled in as further exploration takes place. We can be wrong about some of those properties – just look at the early maps of America, for example – (. . .) but we are rarely wrong about all of them. (*ibid*., p. 25)

We have neglected a central point: the content of any image of the studied paper can be seen as determined by the causal relations involved in producing the data. Images cannot be understood only as symbols standing for something else. They are objects that have a causal relationship to the thing under study. Causal relations are relevant to understanding their role as evidence. Can the concept of robustness be fully developed without thinking about that point? This is a topic for another study.

7.8 Conclusion

As one can see, the argumentation in the astrophysics paper under study takes the form of a robustness analysis, in Wimsatt's sense: multiple derivations are mobilized to establish if the results are robust or not. In our example, the results are the two big maps. It is astonishing to consider as we did, at the same time, images as results and as arguments. In fact, they can have both functions depending on the role they play in the argumentation. For example, the two big maps of the article mentioned above are the results of this publication, and of course, as results, they must be, if possible, very robust. In order to increase this robustness, the scientists use a robustness analysis by exploiting other images published in other papers. But we can notice that if in the present argumentation these images are used as arguments, in the original papers where they come from, the same images were considered as results! One of the difficulties encountered by the researchers is precisely to find the best way to compare the data obtained with different detectors. This is done through data processing methods that can be extremely problematic. The arguments consist in the association of several images whose meaning, context and production procedures are stabilized by the text (caption or main text). Recognizing the similarities between the images lead to the conclusion that results independently produced converge.What the scientist aims to produce is a kind of Peircean cable constituted with many fibers (each fiber is an argument). A cable is robust because it is made of many fibers and, unlike a chain, the solidity of which depends on its weakest link, the cable remains robust even if one or two of its fibers break. (Callebaut 1993, p. 57).

To close this paper we would like to make some brief remarks about the relation between robustness analysis as we have characterized it through our example, and the realist/constructionist issue. If we remain at an epistemological level, our analysis could fit with the constructionist or the realist point of view. But if we consider the ontological level, it is certain that the astrophysicists use the robustness procedures in order to claim the existence of the entities/properties they studied. As philosophers, are we forced to endorse the point of view of the astrophysicists? Nothing is less certain. But nevertheless, as philosophers we have to take into consideration that the concept of robustness analysis (which is in fact a philosophical concept) gives a perfect account of the procedures used in the day-to-day activities of a lab to prove scientific assertions. That is what we wanted to show; the convergence here between the philosophical concept and concrete scientific practices. With the robustness concept, we have a perfect example of a "working" concept that can build a bridge between the scientists and the philosophers. We need that kind of "working" concept if we want to progress in our philosophical investigations of scientific practices.

References

Allamel-Raffin, C. 2004. "La production et les fonctions des images en physique des matériaux et en astrophysique." PhD diss., University of Strasbourg.

Allamel-Raffin, C. 2005. "De l'intersubjectivité à l'interinstrumentalité. L'exemple de la physique des surfaces." *Philosophia Scientiae* 9(1):3–31.

Barthes, R. 1964. *Eléments de Sémiologie*. Paris: Gonthier Médiations.

Callebaut, W. 1993. *Taking the Naturalistic Turn or How Real Philosophy of Science is Done*. Chicago: University of Chicago Press.

Chang, H. 2001. "Spirit Air, and the Quicksilver: The Search for the Real Temperature." *Historical Studies in the Physical and Biological Sciences* 31:260–84.

Collins, H.M. 1981. "Son of Seven Sexes: The Social Destruction of a Physical Phenomenon." *Social Studies of Science* 11(1):33–62.

Collins, H.M. 1992. *Changing Order. Replication and Induction in Scientific Practice*. London: Sage.

Collins, H.M., and T. Pinch. 1993. *The Golem*. Cambridge: Cambridge University Press.

Culp, S. 1995. "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles." *Philosophy of Science* 62:430–50.

Dame, T.M., D. Hartmann, and P. Thaddeus. 2001. "The Milky Way in Molecular Clouds: A New Complete CO Survey." *The Astrophysical Journal* 547:792–813.

Daston, L., and P. Galison. 1992. "The Image of Objectivity." *Representation* 40:81–128.
Daston, L., and P. Galison. 2007. *Objectivity*. New York: Zone Books.
Godin, B., and Y. Gingras. 2002. "The Experimenter's Regress: From Skepticism to Argumentation." *Studies in History and Philosophy of Science* 3(1):137–52.
Hacking, I. 1981. "Do We See Through a Microscope?" *Pacific Philosophical Quarterly* 62:305–22.
Hacking, I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.
Hacking, I. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, dir., A. Pickering, 29–64. Chicago: University of Chicago Press.
Heyer, M.H., C. Brunt, R.L. Snell, J.E. Howe, F.P. Schloerb, and J.M. Carpenter. 1998. "The Five College Radio Astronomy Observatory CO Survey of the Outer Galaxy." *The Astrophysical Journal* 115:241–58.
Hudson, R.G. 1999. "Mesosomes: A Study in the Nature of Experimental Reasoning." *Philosophy of Science* 66:289–309.
Humphreys, P. 2006. *Extending Ourselves: Computational Science, Empiricism and Scientific Method*. Oxford: Oxford University Press.
Jacobi, D. 1985. "La visualisation des concepts dans la vulgarisation scientifique." *Culture Technique* 14:152–63.
Janik, A., R. Rieke, and S.E. Toulmin. 1984. *An Introduction to Reasoning*. New York: Macmillan.
Kitcher, P. 2001. *Science, Truth and Democracy*. Oxford: Oxford University Press.
Kosso, P. 1989. *Observation and Observability in the Physical Sciences*. Dordrecht: Kluwer.
Nederbragt, H. 2003. "Strategies to Improve the Reliability of a Theory: The Experiment of Bacterial Invasion into Cultured Epithelial Cells." *Studies in History and Philosophy of Biological and Biomedical Sciences* 34:593–614.
Putnam, H. 2003. "Pragmatisme et Connaissance Scientifique." In *Cent ans de Philosophie Américaine*, edited by J.-P. Cometti and C. Tiercelin, 135–55. Pau: Pup.
Toulmin, S.E. 1958. *The Uses of Argument*. Cambridge: Cambridge University Press.
Tufte, E.R. 1990. *Envisioning Information*. Cheshire, CT: Graphic Press.
Tufte, E.R. 1997. *Visual Explanations. Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphic Press.
Tufte, E.R. 2007. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphic Press.
Waismann, F. 1945. "Verifiability." *Proceedings of the Aristotelian Society* 19:119–50.
Wimsatt, W.C. 1981. "Robustness, Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins, 124–63. San Francisco, CA: Jossey-Bass Publishers.
Wimsatt, W.C. 1990. "Taming the Dimensions – Visualisations in Science." *PSA* 2:111–38.
Wimsatt, W.C. 2007. *Re-engineering Philosophy for Limited Beings*. Cambridge: Harvard University Press.
Winch, P. 2007. *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge.

Chapter 8
Are We Still Babylonians? The Structure of the Foundations of Mathematics from a Wimsattian Perspective

Ralf Krömer

8.1 Introduction[1]

Wimsatt in his seminal paper on robustness[2] devoted a paragraph to the study of the role of robustness in the structure of theories. In this study, he relied on Feynman's distinction between "Euclidean" and "Babylonian" structure of theories (Feynman 1965). Here, a theory is said to have a Euclidean structure roughly if it is presented axiomatically, with some basic propositions from which the other propositions of the theory can be deduced by serial processes; it is said to have a Babylonian structure if there are no propositions considered as basic, but if the propositions rather form an overconnected system where the various pieces admit multiple determinations by parallel processes, and one can start from various points.[3] Wimsatt applied the statistical theory of reliability to the two types of structure, thus showing that a

[1] I wish to thank my colleagues Amirouche Moktefi, Léna Soler and Emiliano Trizio for commenting in detail on this chapter and helping me considerably improve it and clarifying the main points. Several passages in the chapter have been taken more or less verbally from this discussion. More generally, I am very grateful for the opportunity of collective work inside the PratiScienS group, and for the patience and the valuable help.

[2] Throughout this chapter, I will feel free to make use of the ideas contained in Wimsatt (1981), and paraphrase them without giving always precise quotations. This practice is justified by observing that the reader is supposed on virtually every page of the present volume to have a good knowledge of chapter 2.

[3] See Wimsatt (1981), p. 129 (see Chapter 2); I will adopt the terminology throughout the chapter. I decided to stick with this terminology mostly because I find it charming, even though one could argue that labelling these two types of structure of a theory "Euclidean" and "Babylonian", respectively, is not entirely satisfactory in that it seems to draw upon a historical context which isn't really existing. It is historically justified to label the first type "Euclidean", but this usage can conflict with other established usages of the term when axiomatization of geometry is concerned; choosing the term "Babylonian" for the other alternative seems historically not very motivated. One could simply say "over-connected" for what is meant by "Babylonian".

R. Krömer (✉)
University of Siegen, Siegen, Germany
e-mail: kroemer@mathematik.uni-siegen.de

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_8,

theory with a Babylonian structure is more reliable (or less vulnerable) than one with a Euclidean structure. Wimsatt's analysis concerns the practice of science, in the sense that he doesn't ask what an ideal infallible scientist could do in principle, but what real fallible scientists actually do, and how they can minimize the total error probability.

This amounts to a certain notion of the vulnerability of a theory. A theory is vulnerable by errors committed by the fallible scientists. The vulnerability of mathematical theories discussed in the present chapter is of a different kind: a theory is vulnerable by the discovery of formerly unknown inconsistencies. This second notion of vulnerability obviously doesn't admit of any probability analysis. Nevertheless, as is shown in the chapter, scientists seem to rely on a Wimsattian line of argument concerning the reliability of their theories also with respect to the second notion of vulnerability.

More generally, it is the aim of the chapter to analyse the ways mathematicians organize the structure of their theories and look at their reliability. Here, we might think of a theory in mathematics naively as some collection of mathematical propositions, all concerning some "field" of mathematics. As a rule, in modern mathematics such a theory is presented with a Euclidean (namely axiomatic) structure: in various fields of mathematics, one singled out, following the model of Euclid's elements, some assumptions as "basic" for the theory of the field and tried to deduce the other propositions from them.

But the discovery of various results concerning axiomatics has forced mathematicians to think about the consequences of this standard approach. First of all, some propositions turned out to be logically independent of (i.e. not deducable from) the assumptions considered as "basic" for the theory of the field concerned. One way out of this situation is to enlarge one's axiom system.

But then (and especially with the discovery of several possible enlargements contradictory one with another), the next question came to the fore, namely whether a given axiom system is consistent (or, if it is not, leads to a contradiction and thus to a "total collapse", in Wimsatt's terms).[4]

And now, mathematicians faced the situation that on the one hand, the proposition expressing the consistency of the axiom system can be spelled out in the vocabulary of the field in question and that on the other hand, one can show that this proposition can't be deduced from the axioms. As is well known, this was the case in axiomatic set theory.

The next section of this chapter contains some general remarks on the role of proof (and of the search for multiple proofs) in modern mathematics. Sections 8.3, 8.4, 8.5, and 8.6 present evidence for the presence of Wimsattian perspectives in the debate on the foundations of mathematics of the last two centuries, running from the foundations of geometry, set theory, and category theory, respectively, through the status of Grothendieck's proof techniques. In Section 8.3, I uncover some elements of "Babylonian" methodology in the history of modern axiomatics.

[4] See Wimsatt (1981), p. 134 (see Chapter 2).

In Section 8.4, some utterings of mathematicians are presented which can be read as making appeal to something like Wimsattian robustness in the case of consistency of set theory, and these utterings are submitted to a thorough discussion. In Section 8.5, I present some other utterings which seem to suggest that at least in the case of one mathematical theory (namely category theory), one rather comes back to a "Babylonian" structure.[5] Section 8.6 contains some remarks on the vulnerability (in both the Wimsattian and the logical senses) of famous proofs using category theory. In conclusion, robustness and Babylonian structure turn out to be of some relevance in the ways mathematicians look at the reliability of their theories, but these concepts are not always used in a way appropriate to actually enhance reliability.

In this case study, two propositions which are logically independent of the relevant base of deduction will repeatedly be encountered (and thus deserve to be denoted in abbreviated form):

1. the proposition P_1 asserting the consistency of Zermelo-Fraenkel set theory (base of deduction: ZFC)
2. the proposition P_2 asserting the relative consistency of a certain large cardinal axiom called Tarski's axiom (base of deduction: ZFC + P_1).

The role played by P_2 in the context of category theory will be explained below. Still, this example will look quite technical, but on the other hand it will turn out that the discussion related to it is particularly useful for judging the relevance of "Babylonian" structure and robustness for mathematics.

8.2 Some Remarks on the Role of Proof in Mathematics

In mathematics, the only way to establish a result is proof (in the sense this term is usually employed in mathematical discourse, namely meaning deductive proof establishing the validity of the result once and for all). Thus, there seems not to be much room for multiple determination. But there are at least two problems with this opinion:

1. Proofs can be false, and errors can pass undetected for a long time. This is one of the reasons why confidence in a proof can be larger or smaller, such that reliability theory is an issue.
2. Call a mathematical statement *important* if it has many applications to other mathematical problems or if it affects the structure of various mathematical theories or both. In this sense, many important mathematical statements are still

[5] Category theory is a mathematical discipline which has proved very useful in some of the 'central' mathematical disciplines like topology, algebraic and differential geometry, number theory and functional analysis. The few things readers of the present chapter need to know about category theory will be explained below.

> unproven. Now, it makes a difference in which of the two following situations we are with respect to such an unproven statement. As long as the only thing we know about it is that it is unproven, we will not think about multiple determination in the sense of the empirical sciences as a possible substitute for proof (even if such multiple determination could make the statement more "plausible"). But we can happen to know more about the statement. The statement might be known to be logically independent from the relevant base of deduction. In this situation, there is no point in continuing to look for a proof in that framework; rather, we know that we will never have such a proof (at least if we maintain our conviction that the base of deduction considered is the "relevant" one) and we need to take our decision on other grounds than proof.

Let us discuss the first point first. Following Wimsatt, a proposition of empirical science is *robust* if it admits of a *large* number of *independent* and *converging* confirmations (where one has to make clear in the particular context what it means for a number to be large and for confirmations to be independent).

According to the mainstream view of the epistemology of mathematical proof, this concept of robustness is not useful to mathematics, for as soon as you have one proof, it doesn't matter to have fifty. A proof is seen as something stronger than a mere 'confirmation' in empirical science.[6]

But it is well known that "since ancient times the presentation of new proofs of previously established results has (. . .) been an esteemed and commonplace mathematical practice" (Dawson 2006, p. 269). Whence the question why mathematicians do re-prove theorems. This question has found an exhaustive treatment by John Dawson. Dawson in particular enumerates a variety of reasons to reprove theorems; his last reason is the following:

> (. . .) the existence of multiple proofs of theorems serves an overarching purpose that is often overlooked, one that is analogous to the role of *confirmation* in the natural sciences. For just as agreement among the results of different experiments heightens credence in scientific hypotheses (and so also in the larger theories within which those hypotheses are framed) different proofs of theorems bolster confidence not only in the particular results that are proved, but in the overall structure and coherence of mathematics itself. (p. 281; Dawson's emphasis)

And, while not citing Wimsatt, Dawson goes on by paraphrasing the very passage from Peirce on chains and cables which opens Wimsatt's chapter.[7]

---

[6] When I call this the mainstream view of the epistemology of mathematical proof, I do not make any effort to underpin the implicit claim that this is really the view of some majority; I don't even discuss which group of people whose individuals hold competing views about mathematical proof is relevant to such a statement of majority. I just have the impression that this view is widespread.

[7] See Wimsatt (1981), p. 124 (see Chapter 2). Note that when Dawson speaks about "different" experiments and proofs, this is close to what Wimsatt labels "independent" determinations, as is clear from Dawson's discussion of "criteria for differentiating among proofs" (p. 272ff). Moreover, let me stress again that the quoted paragraph contains only one reason for multiple proofs among several which Dawson listed.

The bolstering of confidence described by Dawson is justified, for as Wimsatt has pointed out in his study of failure probabilities of derivations, “with independent alternative ways of deriving a result, the result is always surer than its weakest derivation” (Wimsatt 1981, p. 132 see Chapter 2, p. 66). As soon as error is possible, it is better to have more confirmations.

Summing up, mathematical proofs seem to play the same role as confirmations in the sense used in the empirical sciences here, and re-proving theorems amounts to giving a Babylonian structure to a theory. From these reflections, we might (and should) be led to feel the need for a more subtle, and more appropriate, epistemological conception of mathematical proof and its role for conviction (eventually making use of the concept of robustness).

In the next three sections of the chapter, I will however step aside from the discussion of cases where the need for further proofs for an already proven proposition has been felt. I rather focus on the case where several confirmations for an unproved proposition have been proposed. I intend to show that Babylonian structure plays an important role in the theories concerned in this second case.[8]

8.3 Logical Independence and the History of Geometry

Before starting the discussion of P_1 and P_2, note that what the two propositions have in common is that they are not only useful (or *would* be useful if indeed they *were* known to be valid propositions) but moreover are logically independent from (that is, outside the deductive scope of) the respective deductive systems considered as relevant for their deduction. In analysing the behaviour of mathematicians with respect to this situation, it will prove useful to have a look at the history of the concept of logical independence in mathematics. We will see that a possible reaction on the discovery of the logical independence of some proposition with respect to its relevant base of deduction was to abandon the idea of that proposition’s having a definite truth value. While in my opinion this would certainly be the right attitude in the case of set-theoretical statements like Tarski’s axiom (see Section 8.5 below), I will argue that this is not the case with proof-theoretical statements like P_1 or P_2 (and we will see in Section 8.4 that this actually wasn’t the reaction in the case of P_1). Another outcome of the historical discussion will be that the feature of logical independence has been turned into a possibility for reducing vulnerability inside axiomatics, hence inside the Euclidean structure of the corresponding theory.

It took some time in the history of mathematics until mathematicians realized that there *is* something like logical independence. The crucial historical example is the axiom of parallels in Euclidean geometry. This axiom, after enormous efforts to

[8] Let me point also to a paper by McLarty treating a particularly interesting example of the search for an alternative proof (McLarty 2010). I shall briefly come back on the issue of this paper in Section 8.6; moreover, I made some hints concerning the problem of multiple proofs in my book (Krömer 2007), especially concerning Lefschetz’s fixed point theorem (in this case, see p. 19 and the sections indicated there).

prove or to refute it from the other axioms, *turned out* to be independent of them. This event challenged the conception of geometry as a unique science of real, physical space. Mathematics hadn't any longer the purpose to make true assertions about empirically given matters.

Moreover, the discovery of the independence of the axiom of parallels motivated a new methodological focus of looking for independence of axiom systems in Hilbertian axiomatics and some of its derivatives: an axiom system (for geometry or some other part of mathematics) was submitted not only to the demands to be complete and consistent, but the axioms had to be mutually logically independent as well. It is well known how Hilbert used this criterion in his *Grundlagen der Geometrie* in order to pass instantly to a different geometry by exchanging particular axioms. More elaborate (and, we might say, nearly perverse) uses of it can be found in the efforts towards an axiomatiziation of mechanics by Hamel and Schimmack (Moore 1995, pp. 273–275), and in early axiomatizations of linear algebra.[9] These developments soon ran into a dead end, and logical independence is not any longer a real issue in usual axiomatizations in modern mathematics. But nevertheless, the phenomenon deserves some words of interpretation.

The idea to look for logical independence in axiomatics presumably was rooted in the very idea of axiomatizing (and thus was related to the project of giving mathematics a Euclidean structure). For to axiomatize means to go down (in the deductive chain)[10] to propositions which can't be proved but have to be taken for granted. Since a proposition is called dependent of others if it can be proved from them, it is clear that such a proposition is a bad candidate for an axiom – whence the search for independence of the axioms.

But the criterion – maybe accidentally, maybe not – also reduces the vulnerability of an axiom system, much like a Babylonian structure reduces the vulnerability of a theory. For by rendering one's axiom system independent, one limits the damage possibly produced by the occurrence of contradictions. If one axiom turns out to be problematic, this does not necessarily harm the other axioms, provided they are independent of it. But recall that we are concerned here with a different notion of vulnerability, compared to Wimsatt's analysis. So it would be a bit misleading to say that we uncovered a feature of Hilbert's axiomatics which might be said to make up a Babylonian structure in mathematics. Rather, the feature of logical independence here has been turned into a possibility for reducing vulnerability inside axiomatics, hence inside the Euclidean structure of the corresponding theory.

The historical impact of logical independence on the foundations of mathematics is not limited to the case of the axiom of parallels. The 20th century saw the famous independence results of Gödel and Cohen (including the consistency of set theory, the axiom of choice, and the continuum hypothesis). At this stage, we can refine

[9] See Krömer (2003) for references.

[10] Wimsatt stressed that the metaphor of a chain is misleading in the context of deduction of propositions, and I think he did this convincingly. It is in the absence of a better terminology that I stick with the metaphor.

our analysis of the different roles multiple determination could play in the case of conjectures and of logically independent statements, respectively. There are actually two types of logically independent statements. There is a fundamental difference between the axiom of parallels on the one hand and the statement of the consistency of axiomatic set theory on the other. Since the discovery of noneuclidean geometries, the axiom of parallels is not any longer supposed to have a definite truth value. Its truth value depends on the meaning which is given to the words point, line etc. in the geometric model chosen. An average mathematician will however believe, I think, that the statement P_1 which states the consistency of axiomatic set theory *has* a definite truth value, even if we don't know this value and never will know it, provided it is "true".[11] For the axiom system is an artefact completely at our disposal,[12] and the rules of inference which can be used in some deduction are equally definite, and we are convinced that either in this formal system one can infer a contradiction or not, and that this property of the formal system is independent of which meaning is given, by some model, to the $\in$-relation.

And I think that it is for this reason that mathematicians tend, as we will see, to make appeal to some idea like multiple determination in the case of P_1 and not (any longer) in the case of the axiom of parallels.

8.4 Bourbaki and the Robustness of the Consistency of Set Theory

Given that consistency of mathematical axiom systems is often unprovable, robustness of the claim of consistency could be a valuable alternative to formal proofs of consistency as a criterion for the acceptability of such an axiom system. According to Bourbaki, this robustness provides empirical instead of mathematically conclusive evidence and thus another kind of conviction (having another source than a formal proof). This opinion, and in particular the sense in which Bourbaki speaks about empirical facts here, will be inspected more thoroughly in what follows.

Let us inspect the case of P_1, the statement of the consistency of ZF. It follows from Gödel's classical paper (Gödel 1931) that P_1 cannot be decided inside ZF. In particular, ZF could be inconsistent; but this could only be proved by finding one day a contradiction.

Now, if there is no proof, are there other ways to increase confidence in P_1?[13] Here is what seems to be Bourbaki's reaction[14]: if there is no consistency proof for a

[11] Maybe an intuitionist will not accord P_1 a definite truth value for this reason.

[12] Here, the case of the axiom scheme of replacement is surely debatable.

[13] In the view of Kreisel, expressed on p. 110 of the German version of his article on the formalist-positivist doctrine (Kreisel 1974), Zermelo provided a compilation of evidence for the consistency of ZF (Zermelo 1930).

[14] I express myself rather carefully here since Leo Corry once challenged me with the question whether there really is such a thing like a philosophical position of Bourbaki. On the one hand,

system, it is considered as "secure" if it has been applied over and over again without producing a contradiction; when problems occur, one looks for ad hoc solutions.

This position is explicitly adopted in the 1954 introduction of *Ensembles* (my translation):

> In the 40 years since one has formulated with sufficient precision the axioms of [set theory] and has drawn consequences from them in the most various domains of mathematics, one never met with contradiction, and one has the right to hope that there never will be one. (...) We do not pretend that this opinion rests on something else than experience. (Bourbaki 1954)

as well as in the talk *Foundations of mathematics for the working mathematician* delivered by André Weil:

> Absence of contradiction, in mathematics as a whole or in any given branch of it, (...) appears as an empirical fact, rather than as a metaphysical principle. The more a given branch has been developed, the less likely it becomes that contradictions may be met with in its further development. (Bourbaki 1949)

I think that the argument is to be criticised for several reasons. In order to express my criticisms, I shall try to elaborate the argument a bit. In my opinion, it rests on the fact that the use of the axioms in the various domains didn't (visibly) lead to a contradiction so far. We can then interpret each such use as a certain *test* of P_1, and we might think of the (alleged) multiplicity of such tests as independent confirmations of P_1. In view of the terminology (initiated by Wimsatt) of calling a result robust if there are multiple independent confirmations, P_1 could hence be seen as robust. More precisely, the more of these confirmations we have, the greater is the probability of the correctness (or the degree of robustness) of our assertion. (It goes without saying that Bourbaki doesn't explicitly refer to Wimsatt, but their trust in their argument seems to rest implicitly on a similar conception.) Clearly, Bourbaki doesn't think of *replacing* the mathematical proof by these multiple confirmations; rather, they consider this multitude as a *sufficient justification in practice*.

Now, my criticisms of this argument concern the following points:

1. A consideration like Wimsatt's is only applicable in situations where parallel processes result in a reduction of a total failure probability. It is not clear to me which kind of probability analysis would be possible in the present case.

Bourbaki has several published accounts on his (or their) position; besides the two cited below, I think of the article "L'Architecture des mathématiques", presumably by Dieudonné (Bourbaki 1948), the article "Les méthodes axiomatiques modernes et les fondements des mathématiques" (Dieudonné 1939) (see also (Dieudonné 1951)) and the article "Sur le fondement logique des mathématiques" (Cartan 1943). On the other hand, Bourbaki is a collective, and no member has exactly the same opinions as the others (the most striking example among the founding members is probably Chevalley); and worse, one can raise doubts about whether Bourbaki took serious the philosophical problems of mathematics at all, and even whether they took serious their own solution to them; finally, it is doubtful that the various ingredients of this solution (structuralism, empirical consistency) are coherent, but this is not a question to be attacked in the present chapter.

2. Moreover, Wimsatt's conclusion rests on the independence of the multiple confirmations – and such an independence is at least debatable in the present case.
3. Bourbaki seems to suggest a certain way in which contradictions are met with in the everyday work of mathematicians, but I am not convinced that this is indeed the usual way in which they have been met with, at least historically.

Let me elaborate each of these criticisms a bit. As to the first issue, I think that we are confronted here with a notion of the vulnerability of a theory different from the one used by Wimsatt. And I actually don't see how to get a clue for any serious probability analysis in this situation.

Weil speaks about the likelihood of the future discovery of contradictions in a branch of mathematics, and about this likelihood's being antiproportional (in some sense) to the rate of development of the branch. His statement would require some preliminary discussion in order to be made precise: how can the rate of development of a branch of mathematics be measured? Weil apparently does think of such a branch as something with a finite potential of development, at least if likelihood means probability in any mathematical sense. For the statement would only be correct if it meant "the closer the development of the branch comes to its end, the less likely etc." But how can one think about the totality of consequences which can be drawn from a set of, e.g., first order axioms as something finite? In the Bourbaki quotation, on the other hand, the process to be analyzed consists in drawing consequences, and failure would mean to meet with contradiction. But what shall we understand by the "probability of meeting a contradiction"?

Now to the second problem: At least in the first quotation, the argument seems to rest not just on the sheer number of consequences drawn from the axioms of set theory without meeting with contradiction, but also on the idea that the "most various domains of mathematics" wherein these consequences have been drawn are independent of each other in some sense. One might debate two things here: in which sense they are indeed independent, and: how weak can the (sense or degree of) independence be in order that the argument still works. Since the first question will turn out to be the hard one, let us first discuss the second question. If we take Wimsatt's idea of parallel processes very verbally, the multiple processes need to be independent with respect to the failure probabilities one investigates, otherwise failure probabilities wouldn't multiply, and we couldn't prove in the way Wimsatt does that the total failure probability reduces. (It is clear at this stage that we should take for granted for the sake of discussion that my first criticism about the very existence of failure probabilities in the present case doesn't apply.) One could say here that there might be subtler probability considerations, allowing partially dependent processes and still arriving at a reduction of the total failure probability. This is a mathematical question which Wimsatt explicitly skipped by corresponding simplifications of his mathematical argument and which I will not discuss here (since I am believing in my first criticism, after all).

Be this as it may, let us now discuss the actual "independence" of the various particular contexts or tests. Bourbaki at least seems to believe (apparently without giving a justification for it) that the different applications of set theory in the different contexts are *sufficiently* different to be considered as "several" (and not just "the same"). But we only have a very naive vision of this independence. For instance, it is questionable proof-theoretically whether "set theory" (as a whole, in the whole strength) has been applied (which means: consequences have been drawn from it) without producing contradictions in a great number of different and separated contexts. For as Georg Kreisel has pointed out, it may very well be that in each particular context one only applies a particular portion of the whole strength of set theory, thus installing a new brand of "independence" amounting to a loss of comparability between these applications.[15]

The problem of independence seems to be even more serious than that. When we speak about the drawing of consequences from the axioms of set theory in various branches of mathematics, we are concerned with the use of set theory as an axiomatic foundation of mathematics, i.e. as a Euclidean structure of (the whole of) mathematics. From the viewpoint of those who believe that the various domains of mathematics can all be reduced to these axioms, these domains just can't be independent (since they all depend on their common axiomatic foundation). They are just different branches of the same tree with the same roots. But this picture is too simple-minded. It describes mathematics as composed of some "branches" or "domains" which, while not having been known to be for a while, in fact *are* nothing but special parts of set theory. But we could also think of these domains as having been there (in many respects independent of each other) before set theory came along and was imposed on it all a posteriori.

The problem of the independence of contexts seems to me a difficult one to the solution of which at present I offer no further contribution. The first idea would certainly be to inspect the methodology used in the empirical sciences for deciding the independence of confirmations and to try to transfer appropriate parts of it to our situation, but unfortunately, it doesn't seem that Wimsatt has developed anything like a thorough methodology allowing to make such decisions in a systematic way.[16]

Anyway, there is a third problem with the Bourbaki approach, in my opinion. This problem is rather a historical (or methodological) one. The quotations suggest that the typical way contradictions in mathematics are found is that "they are 'met with'" (so to say by coincidence) when doing "ordinary" mathematics (i.e. not studying questions like the consistency of one's axiom system, but using the axioms in order to study the field of mathematics they're supposed to axiomatize). But look at the ways contradictions in the history of mathematics have actually been arrived at. Recall that there is a difference between a natural and a contrived development (or, as I once called it, a reasonable and a pathological use of concepts (Krömer 2007)). The Russell antinomy was not discovered by doing ordinary mathematics,

[15] See Krömer (2007) Section 6.6 for references. See also McLarty (2010) for a discussion of similar problems occurring in the analysis of A. Wiles' proof of Fermat's last theorem.

[16] Nor does it seem that someone else so far succeeded in doing so. There are a few studies on what independence of confirmations precisely means in the case of empirical sciences, but so far they leave open many questions. See Chapter 1 for a discussion and references.

but by doing reflexive mathematics in the sense of Corry (1996). Compare the following quote by Gödel:

> As far as sets occur and are necessary in mathematics (at least in the mathematics of today, including all of Cantor's set theory), they are sets of integers, or of rational numbers (...), or of real numbers (...), or of functions of real numbers (...), etc.; when theorems about all sets (or the existence of sets) in general are asserted, they can always be interpreted without any difficulty to mean that they hold for sets of integers as well as for sets of real numbers, etc. (...). This concept of set, however, according to which a set is anything obtainable from the integers (or some other well defined objects) by iterated application of the operation "set of", and not something obtained by dividing the totality of all existing things into two categories, has never led to any antinomy whatsoever; that is, the perfectly "naive" and uncritical working with this concept of set has so far proved completely self-consistent. (Gödel 1947, p. 518f)

It seems as if Gödel much like Bourbaki wishes to conclude from this observation that a future discovery of a contradiction in the *natural* domains of mathematical activity becomes less and less likely (where the term "natural" does not refer, *pace* Gödel's views, to some platonist universe of sets but to the products of the typical operations of mathematical practice). But the actual history of the antinomies of naive set theory indicates that this by no means excludes the discovery of contradictions if only one looks at the right place for them! (As we will see, Gödel's observation rather shows that "ordinary" mathematics has a quite Babylonian structure.)

Summing up, we find that Bourbaki's argument for confidence in the consistency of set theory can't claim to have the strength of Wimsatt's argument in favor of parallel processes. Nevertheless, it is a matter of fact[17] that many mathematicians believe in P_1, be it for the reasons given by Bourbaki or for other reasons. Consequently, proofs of relative consistency (proofs that to make such and such assumption is consistent with ZF) became quite important. In the Feynman terminology, this focus clearly is Euclidean: streamline everything to our trusted basis.

What to do, however, if one adopts new axioms which have not yet been subject to any testing and which do not admit a proof of relative consistency? Are there other possibilities to increase our conviction in *these* cases?

8.5 The Foundations of Category Theory: Another Case for Robustness?

Category theory is a rather recent mathematical discipline which started with work by Eilenberg and Mac Lane (1945). Originally introduced in order to resolve some particular conceptual problems in algebraic topology, the theory later

---

[17] Again, I claim this in the absence of statistical evidence, just from the impression I have got from my discussions with mathematicians.

became an important tool for algebraic geometry and other disciplines in the hands of Grothendieck and others.[18]

However, the reader of the present chapter doesn't need to know much about category theory, let alone about the mathematical problems resolved with its help. The naive idea is to consider totalities of sets bearing a certain structure (the 'objects' of a category) together with the functions defined on these sets respecting the structure (the 'arrows' of a category). For example, there is a category of groups and group homomorphisms, one of topological spaces and continuous functions and so on; there is an algebraic operation on such a category, namely the composition of arrows.[19]

This idea is quite simple, and one hardly sees how it can be made into such powerful a tool. But again, we don't need to see that here. What we need to know about is that this naive conception comes with some foundational problems to be resolved before any results about categories can be considered as established. The foundational problems involved essentially are of the type of the set of all sets. For if you want to consider the category of all topological spaces, for instance, you have the problem that every set can be endowed with at least one topology, such that the class of objects of your category would include the (nonexisting) set of all sets.

Obviously, this and similar problems have been recognized very early in the history of category theory, and various solutions have been presented for them. Most of the solutions include some restriction of the 'all' in 'all groups' etc. to the needs of practical purposes. However, even such restrictions necessitate to take for granted some form of the following set-theoretical axiom about large cardinals:

For any set N, there is a set M having the properties

(1) $N \in M$,
(2) $\forall X, Y : X \in M \wedge Y \subset X \Rightarrow Y \in M$,
(3) $\forall X : X \in M \Rightarrow \wp(X) \in M$,
(4) $\forall X : X \subset M \wedge card(X) \neq card(M) \Rightarrow X \in M$.

This axiom has been introduced by Tarski (1938) in the context of the theory of so-called strongly inaccessible cardinals. He proved that $card(M)$ is strongly inaccessible if and only if M has the properties (2)–(4) above. His axiom thus asserts that there are arbitrarily many strongly inaccessible cardinals.

In the foundations of category theory, the axiom often appears in the form of Grothendieck universes. These are sets having a definite place in the hierarchy of sets, but having strong closure properties with respect to some basic operations of forming new sets from the subsets of the universe. The cardinality of such a universe turns out to be strongly inaccessible, whence the connection with Tarski's

[18] See Krömer (2007) for a detailed analysis of the history of category theory, and for more details on the problems discussed in this section.

[19] This picture is naive, for there are categories (models of the formal definition of the notion of category) whose objects can't be seen as structured sets and whose arrows can't be seen as functions on them. But the picture serves the present introductory purpose.

axiom. The postulation of infinitely many universes doesn't allow to speak about the category of *all* groups, but for every group whatsoever, an appropriate universe (and thus an appropriate category of groups in the universe) can be found. The device of Grothendieck universes is by now the most often used set-theoretical foundation for category theory.

Jean Bénabou discussed criteria for the acceptability of a solution of the foundational problems of category theory. In particular, in order to be acceptable,

> "foundations" (...) for category theory [should be] consistent, or at least relatively consistent with a well-established and 'safe' theory, *e.g.* (...) ZF. (Bénabou 1985, p. 10)

(Note that Bénabou relies on the same argument as Bourbaki here.)

How about this in the case of Grothendieck universes? Can Tarski's axiom be deduced from ZF, or can one at least prove its relative consistency with ZF? As to the first point, it turns out that the axiom is independent of ZF; more precisely, it is consistent with ZF to assume that there is only one inaccessible cardinal (namely ω) (Drake 1974, p. 67). (This comes as no great surprise since the axiom actually has been introduced as a strengthening of ZF, ZF being inappropriate both for the theory of strongly inaccessible cardinals and the needs of category theory.) But what about the second point? Let us call P_2 the statement of relative consistency of Tarski's axiom with ZF; then P_2 is undecidable (Kunen 1980, p. 145).

This fact was not yet known to early workers in the field. In 1965, relative consistency was "*suspected with conviction*" (Kruse 1965, p. 96). In an appendix to SGA 4 (Artin et al. 1972) concerning universes (authored by Nicolas Bourbaki), we read (my translation):

> it would be quite interesting to show that the axiom (...) of universes is not offensive. This seems difficult, and it is even unprovable, Paul Cohen says. (SGA 4 *exposé* I p. 214)

It is possible that Bourbaki encountered the problem already in the late 1950s when discussing foundations of category theory in view of writing a chapter on the topic in the *Éléments de mathématique*. The undecidability result (or rather, from Bourbaki's perspective in the late 1950s, the fact that the question is open) might very well have played a role in Bourbaki's rejection of categories; however, no explicit evidence for this has been found so far in the sources covering the Bourbaki discussion.[20]

Let us come back to our discussion of robustness in mathematics. We see now why the example of P_2 is worth the technical trouble: because it concerns a rather young mathematical theory.[21] Bourbaki's claim (that once a discipline has reached a certain state of development, the future discovery of contradictions becomes unlikely) is not useful in such a case. The last remaining "warrant" is perhaps that the axiom was adopted precisely to *avoid* (known) contradictions arising from naive category theory.

[20] See Krömer (2006).

[21] Category theorists already trusted in P_2 when it still was quite young, namely in the late 1950s.

Here are two more utterings from workers in the field which help us to analyze the source of confidence at issue here.

> The well known fact that some basic constructions applied to large categories take us out of the universe seems to me to indicate that the constructions are not yet properly presented. The discovery of proper presentations is too difficult, though, for all work on these constructions to wait for it. (Isbell 1966)

> Categoricians have, in their everyday work, a clear view of what could lead to contradiction, and know how to build ad hoc safeguards. (Bénabou 1985)

These quotations suggest that the trust of the workers in the field is not so much in P_2 in the absence of proof, but in category theory in the absence of properly presented foundations (P_2 being part of an unproper presentation). The search for a Euclidean structure of category theory is postponed in favour of a Babylonian structure, as it comes along with the everyday work on the problems to the solution of which category theory has proved helpful. Let us thus inspect this helpfulness more closely.

8.6 Grothendieck's Algebraic Geometry, Fermat's Last Theorem and "Theoretical Organization"

Grothendieck's rewriting from scratch of algebraic geometry (making essential use of concepts and results of category theory) is an ingredient of the most important progress made in number theory in the last 30 years (conjectures of Weil, Mordell, Fermat...). It is not clear whether this use of category theory is indispensable; up to now, no alternative proof of Fermat's last theorem is known, for instance. McLarty discusses in detail the proof-theoretical aspects of this situation, the directions in which alternative proofs are currently searched for and the epistemological significance they could possibly have (McLarty 2010). McLarty in particular stresses that the strength of the existing proofs is not that they use minimal assumptions from a logical point of view, but to the contrary that they attack the problem in a highly developed conceptual framework, thus being convincing by what McLarty calls "theoretical organization". It is worthwhile to investigate how these proofs manage to be convincing, given the fact that they use the controversial assumptions discussed in the last section (and thus just *aren't* proofs in the standard Euclidean framework based on ZF). Did the workers in the field just switch to another set of basic assumptions, giving rise to a Euclidean framework based on something else than ZF? Or has the high level of development of the conceptual framework, the "theoretical organization", the effect of giving a Babylonian rather than Euclidean structure to the theory?

Note first that in Grothendieck's practice, the search for theoretical organization is more important than the actual working out of the proofs. Here is what Bénabou says about Grothendieck's work on fibered categories in SGA 1:

> The proofs are long and tedious, but straightforward verifications, mostly left to the reader because they would add nothing to our understanding of fibrations, and moreover one is convinced from the beginning that the result has to be true. (Bénabou 1985)

Once again: how can the "theoretical organization" provide this conviction? The structure of the theory apparently is not Euclidean, at least not in every respect, since according to Bénabou's statement, the insight into the truth of a result (and actually, the understanding of the corresponding mathematical objects) is not achieved by a decomposition of a line of deduction into elementary steps (Descartes' chains dismissed, to take up Wimsatt's picture).

On the other hand, what seems to be claimed is that this insight and this understanding are achieved by transition to appropriate levels of synthesis. You have to know prohibitively many things about Grothendieck's conceptual framework (many of them having nothing to do with category theory proper, but coming from other parts of mathematics like Galois theory, topology and so on) in order to appreciate the snowslide-like culmination of the proof of such and such famous conjecture after thousands of pages. Again, to check such a proof is difficult, let alone to produce alternative proofs using different if not simpler concepts in order to install some Babylonian structure in the theory. The conceptual framework is highly organized and thus *inter*connected, but this does not imply that the theory is *over*connected in Wimsatt's sense.

To the contrary, such a highly organized and interconnected framework is likely to give rise to a rather vulnerable theory, both in the logical and in Wimsatt's sense of vulnerability: if the various pieces of the theory, instead of being kept independent of each other in some respect, are all unified by the same conceptual apparatus, occurring problems are not likely to stay local; if the scientists need an exceedingly huge amount of expert knowledge to master the theory, they are likely to commit errors.[22] (No trace of Peirce's cables either.)

From a foundationalist point of view (favoring Euclidean structure), it is the fact that category theory has in the last analysis unresolved foundational problems which makes fragile the known proofs of the conjectures of Weil, Mordell, and Fermat. In the last section, I discussed the attitude of workers in the field towards these problems; this discussion suggested that they try to give category theory a Babylonian rather than a Euclidean structure, which, from a Wimsattian perspective, would make category theory robust, not fragile. But still the results proved with the help of category theory in algebraic geometry seem vulnerable: given the huge expert knowledge needed, error is quite probable, and alternative proofs making the theory as a whole less vulnerable are highly desirable. If mathematicians in the future engage in some systematic search for alternative proofs of these conjectures, then this could be motivated by the desire to increase confidence in the existing proofs, in the sense of Dawson.

---

[22] I am aware of speaking about likelihood here as naively as Weil did; at least in the second place, I feel secured by the fact that I refer to something finite, namely the mental capacity of researchers.

8.7 Conclusion

What is the outcome of our attempt to use a Wimsattian language in the philosophy of mathematics? On the one hand, it seems that this language turned out to be quite appropriate to describe the behaviour of mathematicians in various parts of the recent discussion of the foundations of mathematics; moreover, the case study presented suggested that the structure of mathematical theories sometimes is more Babylonian (and deliberately so) than one would expect. On the other hand, the reader might have got the impression that we arrived at somewhat diverging results in Sections 8.4 and 8.5, respectively.[23] In Section 8.4, I tried to show that Bourbaki gave a Wimsatt style argument in favor of the reliability of some basic assumption in the foundations of mathematics but that their argument failed, while in Section 8.5 I suggested that the effort (visible in my opinion in the debate on the foundations of category theory) to give the theory a Babylonian structure indeed might reduce the vulnerability of this theory (*pace* Section 8.6 where I stressed that still important mathematical progress made with the help of this theory is vulnerable by the absence of alternative proofs). What is the difference of behaviour in the two cases?

I pointed out the reasons for which Bourbaki's argument fails in my opinion. Actually, part of these reasons could be subsumed under the heading that they tried to do it inside a Euclidean rather than a Babylonian framework. And it is clear that in a Euclidean framework, a Wimsatt style argument just can't apply. In the debate on the foundations of category theory, on the other hand, one deliberately chose to give the theory a Babylonian structure. Still, this has in common with Bourbaki's strategy that one just stops to care about resolving the foundational issues once and for all. But Bénabou's sentence that categoricians "have a clear view of what could lead to contradiction, and know how to build ad hoc safeguards" stresses that mathematical work is no longer supported by pseudo-empirical trust in consistency (of naive category theory in this case), but that one thinks to have learned to live with the problems and how to keep their scope local.

Now, once we have seen that Wimsatt's analysis of reliability is relevant to mathematical practice, we might ask how convincing the lines of argument uncovered in the thinking of the workers in the field are to philosophers. Mathematicians faced with propositions considered as useful but admitting no proof in accepted deductive systems exhibit a behaviour similar to that of experimental scientists. In such a situation, they find multiple confirmations relevant. But they rely mostly on their experience as workers in the field; they don't have a thorough methodology to assure independence of the confirmations. They rely on the robustness of certain propositions, but they can't claim to have checked whether these propositions really *are* robust. Similarly, they rely on the overconnectedness of the structure of their theory, but they can't claim to have checked whether it really *is* overconnected in some relevant sense. Thus, there is work left to be done.

---

[23] I wish thank one of the anonymous referees for having pointed me to that problem.

References

Artin, M., A. Grothendieck, and J. L. Verdier. 1972. *Théorie des topos et cohomologie étale des schémas. Tome 1: Théorie des topos. Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964 (SGA 4)*. 269 *Lecture Notes Math*. Berlin: Springer.

Bénabou, J. 1985. "Fibered Categories and the Foundations of Naive Category Theory." *The Journal of Symbolic Logic* 50(1):10–37.

Bourbaki, N. 1948. "L'Architecture des mathématiques." In *Les grands courants de la pensée mathematique*, edited by F. LeLionnais, 35–47. Paris: Blanchard.

Bourbaki, N. 1949. "Foundations of Mathematics for the Working Mathematician." *The Journal of Symbolic Logic* 14:1–8.

Bourbaki, N. 1954. *Eléments de mathématique. Première partie. Livre I: Théorie des ensembles. Chapitre 1: Description de la mathématique formelle. Chapitre 2: Théorie des ensembles*. Paris: Hermann.

Cartan, H. 1943. "Sur le Fondement Logique Des Mathématiques." *Revue Science (Review of Rose Illustration)* 81:3–11.

Corry, L. 1996. *Modern Algebra and the Rise of Mathematical Structures*. Science Network *Historical Studies*, vol. 17. Basel: Birkhäuser. Zbl.858.01022; MR97i:01023.

Dawson, J.W.J. 2006. "Why do Mathematicians Re-Prove Theorems?" *Philosophia Mathematica* 14(3):269–86.

Dieudonné, J. 1939. "Les Méthodes Axiomatiques Modernes Et Les Fondements Des Mathématiques." *Review of Science* 77:224.

Dieudonné, J. 1951. "L'axiomatique Dans Les Mathématiques Modernes." In *Congrès International de Philosophie des Sciences, Paris, 1949, vol. III, Philosophie Mathématique, Mécanique*, 47–53. Paris: Hermann & Cie.

Drake, F.R. 1974. *Set Theory. An Introduction to Large Cardinals*. Studies in Logic and the *Foundations of Mathematics*, vol. 76. North–Holland.

Eilenberg, S., and S. Mac Lane. 1945. "General Theory of Natural Equivalences." *Transactions of the American Mathematical Society* 58:231–94.

Feynman, R.P. 1965. *The Character of Physical Law*. Cambridge, MA: MIT Press.

Gödel, K. 1931. "Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme." *Monatshefte* 38:173–98.

Gödel, K. 1947. "What is Cantor's Continuum Problem?" *American Mathematical Monthly* 54:515–25.

Isbell, J. 1966. "Structure of Categories." *Bulletin of the American Mathematical Society* 72:619–55.

Kreisel, G. 1974. "Die Formalistisch–Positivistische Doktrin der Mathematischen Präzision im Licht der Erfahrung." In *Mathematiker Über Die Mathematik*, edited by Michael Otte, 64–137. Berlin: Springer.

Krömer, R. 2003. "Axiomes Pour Les Espaces Vectoriels 1918–1923." In *Histoires des Géométries. Textes du Séminaire de L'année 2003*, edited by D. Flament, 119–28. Paris: Maison des Sciences de l'Homme.

Krömer, R. 2006. "La 'Machine De Grothendieck', Se Fonde-T-Elle Seulement Sur Des Vocables Métamathématiques? Bourbaki Et Les Catégories Au Cours Des Années Cinquante." *Revue d'Histoire des Mathématiques* 12:111–54.

Krömer, R. 2007. *Tool and Object. A History and Philosophy of Category Theory*. Science Network *Historical Studies*, vol. 32. Basel: Birkhäuser.

Kruse, A. 1965. "Grothendieck Universes and the Super–Complete Models of Shepherdson." *Compuational Mathematics* 17:96–101.

Kunen, K. 1980. *Set Theory. An Introduction to Independence Proofs*. Studies in Logic and the *Foundations of Mathematics*, vol. 102. Amsterdam: North-Holland.

McLarty, C. 2010. "What does it Take to Prove Fermat's Last Theorem? Grothendieck and the Logic of Number Theory." *Bulletin of Symbolic Logic* 16(3):359–77.

Moore, G.H. 1995. “The Axiomatization of Linear Algebra: 1875–1940.” *History of Mathematics* 22:262–303.

Tarski, A. 1938. “Über Unerreichbare Kardinalzahlen.” *Fundamenta Mathematicae* 30:68–89.

Wimsatt, W.C. 1981. “Robustness, Reliability, and Overdetermination.” In *Scientific Inquiry in the Social Sciences (a festschrift for Donald T. Campbell)* edited by Brewer, M. and Collins, B., 123–62. San Francisco, CA: Jossey-Bass.

Zermelo, E. 1930. “Über Grenzzahlen und Mengenbereiche.” *Fundamenta Mathematicae* 16:29–47.

Chapter 9
Rerum Concordia Discors: Robustness and Discordant Multimodal Evidence

Jacob Stegenga

But to stand in the midst of this rerum concordia discors and of this whole marvelous uncertainty and rich ambiguity of existence…
Nietzsche, Gay Science I.2

A symphony of Beethoven presents to us the greatest confusion, which yet has the most perfect order at its foundation, the most vehement conflict, which is transformed the next moment into the most beautiful concord. It is rerum concordia discors, a true and perfect picture of the nature of the world which rolls on in the boundless maze of innumerable forms…
Schopenhauer, Metaphysics of Music

Quid velit et possit rerum concordia discors. Empedocles deliret acumen?
What does the discordant harmony of things mean, and what can it do? Is Empedocles crazy?
Horace, Epistles I.12.19

9.1 Introduction: Multimodal Evidence

We learn about particular aspects of the world with multiple methods. Galileo's defense of heliocentrism was based on late-sixteenth century astronomical novelties, Brahe's naked-eye observations of Mars and Kepler's accounting of them with elliptical orbits, and Galileo's own telescopic observations of Jupiter's moons and shifting sunspots. Evidence mustered to support Wegener's theory of continental drift included paleontological parallels between continents, stratigraphic parallels between continents, and the jigsaw-puzzle fit of continents. When Avery and his colleagues suggested that genes might be composed of deoxyribonucleic

J. Stegenga (✉)
University of Toronto, Toronto, ON, Canada
e-mail: jacob.stegenga@utoronto.ca

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_9,

acid (DNA), their evidence included chemical analysis, enzymatic experiments, ultraviolet absorption, electrophoresis, and molecular weight measurements.[1] When Tom Ridge was the governor of Pennsylvania he signed the death warrant of Mumia Abu-Jamal, who is accused of shooting a police officer and now sits on death row; Abu-Jamal's purported guilt is supported by testimony of four direct witnesses, and the retrieval of his gun and spent cartridges at the murder scene, which matched the bullets extracted from the murdered officer.

Galileo also had to consider contrary evidence: that bodies fall straight to earth, for example, and evidence of an altogether different kind – the authority of sacred texts – since Ecclesiastes tells us that "the sun also rises." There was evidence against continental drift: for example, data indicated that the Earth's mantle is rigid. Evidence that proteins are the functional basis of genes, rather than DNA, was also manifold: proteins are sufficiently diverse in structure and function to be the basis of heredity, in contrast to the supposed simplicity of DNA in the 1940s, and it was highly probable that Avery's samples of DNA also included undetected contaminating protein. The purported innocence of Abu-Jamal was supported by testimony from multiple witnesses, and after the original guilty verdict (but before Ridge's condemnation) the admission of someone else as the killer, and other information that suggested that much of the original prosecution evidence was flawed.

Galileo, Wegener, Avery, and Ridge relied on "multimodal evidence." Some call this "evidential diversity" (e.g. Fitelson 1996). That term is fine with me. It is a noun. Mine is an adjective – multimodal – and my modified noun is "multimodal evidence". This is a useful neologism because it allows talk of individual modes of evidence and their various relations to each other and to different hypotheses. It is also a salient neologism because it calls to mind our sensory modalities; much sensation is literally multimodal evidence. Locke argued that we are more likely to believe in primary qualities because we observe them with multiple sensory modalities, as when we observe extension with both touch and sight, whereas secondary qualities, like color, we observe only with a single sensory modality. What I mean by "mode" is a particular way of finding out about the world; a type of evidence; a technique or a study design. The total set of evidence that is relevant to a hypothesis of interest and that is generated by multiple modes I call *multimodal evidence*.

When multimodal evidence for a hypothesis is concordant, that is often said to be epistemically valuable.[2] Evidence that is varied is said to provide more support to a hypothesis than does homogeneous evidence. This is how Hempel put it: "The confirmation of a hypothesis depends not only on the quantity of the favorable evidence available, but also on its variety: the greater the variety, the stronger the

[1] See Westman (2011); Oreskes (1999); Stegenga (2011).

[2] Many philosophers of science have claimed that concordant multimodal evidence is useful, including Hempel (1966), Wimsatt (1981), Horwich (1982), Cartwright (1983), Hacking (1983), Franklin and Howson (1984), Howson and Urbach (1989), Trout (1993), Mayo (1996), Achinstein (2001), Staley (2004), Chang (2004), Douglas (2004), Allamel-Raffin (2005), Weber (2005), Bechtel (2006), Kosso (2006). The contributions to this volume are some of the first to critically evaluate such arguments.

resulting support" (1966, p. 34). Conversely, when multimodal evidence is discordant, that is often said to be conducive to uncertainty – there are several responses one hears: the evidence is too messy to know what to believe; or, the 'weight of the evidence' more strongly suggests this hypothesis over that; or, more research is required. Hempel, in the passage above, only mentioned the quantity and variety of "favorable" evidence, but surely confirmation must depend on both favorable and unfavorable evidence. Without a method of systematically assessing and combining multimodal evidence, both views – that concordant multimodal evidence is a Good Thing, and that discordant multimodal evidence is a Bad Thing – are, as I argue in Section 9.6, unsatisfactory.

Multimodal evidence is an exceptionally important notion: it is ubiquitous in science and law; it elicits both certainty and dissent amongst practitioners; and yet it is poorly understood. The above remarks suggest three questions: (1) determining what multimodal evidence *is*; (2) specifying *why* multimodal evidence is valuable; and (3) describing *how* multimodal evidence should be assessed and combined to provide systematic constraint on our belief in a hypothesis. There is little literature addressing the first question; there have been several answers suggested for the second question, one of the most prominent of which is the notion of robustness; and there are several disputed approaches to the third question.

In Section 9.5 I address the first question, and conclude that determining criteria for defining a mode of evidence is a difficult conceptual problem, the solution to which will likely be relative to the way one wishes to use multimodal evidence (this is what I call the *individuation problem* for multimodal evidence). First, though, I discuss two of the prominent answers to the second question: the notion of 'robustness' is one account of how multimodal evidence is said to be valuable (Section 9.2), and 'security' is another (Section 9.3). My explication of multimodal evidence ends in Section 9.6, where I discuss the challenge of assessing and amalgamating multimodal evidence.

9.2 Robustness

One of the primary ways in which multimodal evidence is purported to be valuable is because concordant multimodal evidence is said to be better evidence for a hypothesis, ceteris paribus, than evidence from a single mode; hypotheses supported by concordant independent multimodal evidence are said to be *robust*. Robustness is a recent term that undergirds a common platitude: hypotheses are better supported with plenty of evidence generated by multiple techniques that rely on different background assumptions. A simple example of this was given by Ian Hacking when he argued that if a cellular structure is observed with different types of microscopes, then there is more reason to believe that the structure is real (1983). I have seen the term "robustness" first used as a methodological adage by the statistician George Box in 1953 – a robust statistical analysis is one in which its conclusions are consistent despite changes in underlying analytical assumptions. In philosophy of science I have seen the term first used with respect to models: results consistent across multiple models (with different background assumptions) are 'robust' and so more likely to be true (Levins 1966; Wimsatt 1981); Levins' infamous quip is that "our truth is the intersection of independent lies." Nearly every philosopher of science interested in evidence has, at least in passing, extolled the virtues of robustness.

Thus, robustness can be a feature of statistical analyses, models, and hypotheses. My concern in this chapter is with empirical hypotheses.

Robustness: A hypothesis is robust if and only if it is supported by concordant multimodal evidence.

Another name that concordant multimodal evidence has gone by is "independent determinations" (see, for example, Wimsatt 1981 and Weber 2005). The common presumption is that robustness is epistemically valuable, since concordant multimodal evidence provides greater confirmational support to a hypothesis than does evidence from a single mode of evidence. My definition above has an element of independence between lines of evidence, or 'determinations', built in, since the notion of multimodal evidence is assumed to have a criterion of individuation for modes of evidence. However, as suggested in Sections 9.2 and 9.5, determining both *how* and *if* modes of evidence are independent is difficult. The value of robustness is often simply assumed or left implicit, but one way to understand robustness is as a no-miracles argument: it would be a miracle if concordant multimodal evidence supported a hypothesis and the hypothesis were not true; we do not accept miracles as compelling explanations; thus, when concordant multimodal evidence supports a hypothesis, we have strong grounds to believe that it is true.

Robustness is often presented as an epistemic virtue that helps us achieve objectivity. Champions of robustness claim that concordant multimodal evidence can demarcate artifacts from real entities, counter the "experimenter's regress," ensure appropriate data selection, and resolve evidential discordance. Consider the worry about artifacts: if a new technique shows x, the observation of x might be due to a systematic error of the technique rather than due to the reality of x. *Response*: if x is observed with concordant multimodal evidence it is extremely unlikely that x is an artifact (Hacking 1983). Consider the "experimenter's regress": good evidence is generated from properly functioning techniques, but properly functioning techniques are just those that give good evidence (Collins 1985). *Response*: this vicious experimental circle is broken if we get the same result from concordant multimodal evidence (Culp 1994). Consider the concern about data selection: scientists use only some of their data, selected in various ways for various reasons, and the rest is ignored – but how do we know that the selection process gives true results? *Response*: vary the selection criteria, and invariant results are more likely to be true (Franklin 2002). Finally, consider discordant data: multiple experimental results do not always agree – which results should we believe? *Response*: simply conduct more experiments until they yield concordant multimodal evidence.

Robustness has been used as an argument for realism. The canonical example is Jean Perrin's arguments for the reality of atoms (described in Nye 1972 and discussed in Cartwright 1983; Salmon 1984; and Mayo 1996). Jean Perrin calculated Avogadro's number consistently, using different kinds of experiments:

Brownian motion, alpha particle decay, X-ray diffraction, blackbody radiation, and electrochemistry, and the common-cause for this consistency is the existence of molecules.

Given the variety of epistemic tasks placed on robustness, and given the frequency with which the notion is appealed to, it has received surprisingly little direct philosophical evaluation; the chapters in this volume are an important contribution towards understanding the value and challenges of robustness. I will discuss several problems with robustness, in an attempt to provide needed constraints on the concept. Robustness is valuable in ideal evidential circumstances, when all available evidence is concordant. One major difficulty for robustness is that in many cases multimodal evidence is not concordant. When multimodal evidence is available for a given hypothesis, the evidence is often discordant; that is, evidence from various modes supports competing hypotheses. The general applicability of robustness is mitigated by the problem of discordant evidence. Moreover, scientists have some methods for assessing and combining multimodal evidence, but without using such methods in a robustness-style argument, such an argument is at best a pump of one's intuitions justifying a vague or qualitative conclusion.

9.2.1 Three Preliminary Challenges

Prior to discussing what I consider to be the 'hard' problems of robustness – discordance and individuation – I discuss three preliminary challenges. First, scientists do not always have multiple modes of evidence with which to make a robustness-style argument; second, knowing that multiple modes are independent is difficult or impossible (as is knowing in what way multiple modes should be independent; I discuss this in Section 9.5); and finally, concordant multimodal evidence will not necessarily give a correct conclusion. None of these problems taken alone completely repudiates the value of robustness. Indeed, it is a (trivially) important methodological strategy which scientists frequently use. However, the value of robustness is mitigated, and its extent of application constrained, upon consideration of these three preliminary challenges.

Generating concordant multimodal evidence is difficult. Scientists need evidence from independent modes to make a robustness claim, but they do not always have multiple independent modes of evidence to study the same subject. New modes are introduced into scientific practice for good reason: they give evidence on a new subject, or on a smaller or larger scale, or in a different context, than do existing modes. Even if multiple modes do exist, it is not always clear that they are independent. Bechtel (2006) argued that since new techniques are often calibrated to existing techniques even when both techniques provide concordant results the techniques might fail to be independent (see also Soler's discussion on 'genetic non-independence' in the introduction). Furthermore, determining what criteria should be used to determine independence between modes is a difficult problem; this is what I call the "individuation problem" for multimodal evidence (Section 9.5). Simply put, the following challenges must be met to make a robustness argument:

one must have independent modes of evidence, one must have a criterion to which one can appeal in order to demarcate modes of evidence, and one must know that the available modes meet this criterion so that we can be confident that the modes are properly independent. Since robustness requires multiple modes of evidence, an incomplete or vague individuation of evidential modes will render robustness an incomplete or vague notion, and hence robustness-style arguments will be vague or inconclusive.

One might think that multiple invalid arguments that reach the same conclusion give no better reason to believe this conclusion than a single invalid argument reaching the same conclusion. Similarly, multiple methodologically mediocre experiments, or multiple epistemically unrelated experiments, or multiple modes of evidence with implausible background assumptions, give no better reason to believe a hypothesis than does a single mode (let alone a single well-performed mode with more plausible background assumptions). A detailed case-study discussed by Nicolas Rasmussen provided an instance of this problem: multiple methods of preparing samples for electron microscopy demonstrated the existence of what is now considered an artifact (1993). Although this case study generated a good amount of controversy – see responses from Culp (1994), G. Hudson (1999), and others – the fact that such evidential diversity was used as an argument for the reality of an artifact mitigates the epistemic value of robustness. The problem demonstrated by Rasmussen can be generalized: concordant multimodal evidence can support an incorrect conclusion.

In short, to make a compelling robustness argument, one needs evidence from multiple modes for the same hypothesis, while ensuring that such modes are sufficiently independent. Scientists are often adept at grappling with these challenges. However, the problem raised by Rasmussen indicates that arguments based on robustness can generate incorrect conclusions. In other words, robustness requires having multiple modes of evidence, knowing that multiple modes of evidence are independent and knowing how they should be independent, and yet remains fallible. Knowing *that* multiple modes of evidence are independent depends on knowing *how* multiple modes of evidence must be independent to be sufficient for a robustness argument. The former obviously depends on the latter. In Section 9.5 I discuss the latter problem: what I call the individuation problem for multimodal evidence.

9.3 Security

It is a familiar platitude that data is only evidence with respect to a hypothesis, and to think that data is relevant to a hypothesis we must accept certain background assumptions. The confirmation relation should be construed as a three-place relation between a hypothesis, evidence (from multiple modes), and the various background assumptions required to relate evidence from each mode to the hypothesis. Background assumptions are like any belief: they have varying degrees of plausibility. Some are dodgy. A mode of evidence can provide independent evidence for a background assumption of another mode of evidence. Thus, one evidential mode can support a background assumption which is necessary to relate evidence from another mode to the hypothesis; of course, the evidential support for the first background assumption will require its own background assumptions. Staley (2004) has argued that this is an important use of multimodal evidence. The background assumptions of a single mode of evidence can themselves be supported by independent evidence. Then, when the first mode of evidence confirms a hypothesis, the support that this evidence provides to the hypothesis is indirectly strengthened by evidence from other modes which support auxiliary assumptions required for the first mode.

This is a compelling and rather straightforward way to construe the value of multimodal evidence. We should be clear about the difference between security and robustness. Security does not require multiple concordant modes of evidence for the *same* hypothesis. After all, security just is the use of one mode of evidence to support an auxiliary hypothesis for another mode of evidence, which is itself evidence for the main hypothesis of interest. Thus, security avoids the challenge of amalgamating multimodal evidence which I discuss in Section 9.6. Indeed, one can gain security simply by using a single mode of evidence for a hypothesis, as long as the auxiliary hypotheses for this mode of evidence are supported by other, independent modes of evidence. One might think that we can construe such an evidential situation as robustness with a single mode of evidence. However, it is helpful to maintain the distinction between robustness and security, since the structure of the arguments are different. Moreover, we should not be misled by diction. Security, presumably, is a matter of degree: if the auxiliary hypotheses of a primary hypothesis are supported by independent evidence, then we might be justified in thinking that our primary hypothesis is 'more secure' than if the auxiliary hypotheses were not supported by independent evidence, but we would not be justified in thinking that our primary hypothesis is 'secure' tout court.

9.4 Rerum Concordia Discors

If concordant multimodal evidence provides greater epistemic support to a hypothesis, it is unclear what support is provided to a hypothesis in the more common situation in which multimodal evidence is discordant. Franklin recently raised the problem of discordance, and suggested that it can be solved by various methodological strategies, which prominently include generating more evidence from independent techniques (2002). While Franklin is correct to identify discordance as a problem for what he calls the "epistemology of evidence", and his appeal to a plurality of reasoning strategies is valuable, I argue below that what he considers a solution to the problem of discordance is better construed as the source of problem.

Discordance is based on both inconsistency and incongruity. Inconsistency is straightforward: Petri dishes suggest x and test tubes suggest $\neg x$. In the absence of a methodological meta-standard, there is no obvious way to reconcile various kinds of inconsistent data. Incongruity is even more troublesome. How is it even possible for evidence from different types of experiments to cohere? Evidence from different types of experiments is often written in different ‘languages’. Petri dishes suggest *x*, test tubes suggest *y*, mice suggest *z*, monkeys suggest 0.8*z*, mathematical models suggest 2*z*, clinical experience suggests that sometimes *y* occurs, and human case-control studies suggest *y* while randomized control trials suggest ¬*y*. To consider multimodal evidence as evidence for the same hypothesis requires more or less inferences between evidential modes. The various ‘languages’ of different modes of evidence might be translatable into languages of other modes, if one holds the right background assumptions. That is, seemingly incongruous modes of evidence can both be construed as evidence for the same hypothesis given certain background assumptions that relate each mode to the hypothesis. The background assumptions necessary for such translations will have varying degrees of plausibility. If they are not plausible, then it is hard to see how multimodal evidence provides greater epistemic support to a hypothesis than does a single mode of evidence.

For much of the twentieth century, philosophy of science considered idealizations of evidence – Carnap, for example, developed confirmation theory “given a body of evidence *e*”, without worrying about what constitutes a “body of evidence” (1950). In ideal evidential contexts, robustness is a valuable epistemic guide. Real science is almost never in ideal evidential contexts; recent historical and sociological accounts of science have reminded philosophers of the messy details of scientific inquiry. In Section 9.1 I quickly mentioned Galileo, Wegener, Avery, and Ridge as examples of people grappling with discordant multimodal evidence. The following example more richly illustrates the problem, though the example should hardly be needed, since discordance is ubiquitous.

9.4.1 Multimodal Evidence on Influenza Transmission

Epidemiologists do not know how the influenza virus is transmitted from one person to another. The mode of infectious disease transmission has been traditionally categorized as either “airborne” or “contact”.[3] A causative organism is classified as airborne if it travels on aerosolized particles through the air, often over long distances, from an infected individual to the recipient. A causative organism is classified as contact if it travels on large particles or droplets over short distances and can survive on surfaces for some time. Clinicians tend to believe that influenza is spread only by contact transmission. Years of experience caring for influenza patients and observing the patterns of influenza outbreaks has convinced them that the influenza virus is not spread through the air. If influenza is an airborne virus, then patterns of influenza transmission during outbreaks should show dispersion over large distances, similar to other viruses known to be spread by airborne transmission. Virtually no influenza outbreaks have had such a dispersed pattern of

[3] I am, of course, greatly simplifying for the sake of exposition.

transmission. Moreover, nurses and physicians almost never contract influenza from patients, unless they have provided close care of a patient with influenza.

Conversely, some scientists, usually occupational health experts and academic virologists, believe that influenza could be an airborne virus. Several animal studies have been performed, with mixed conclusions. One prominent case often referred to is based on an airplane that was grounded for several hours, in which a passenger with influenza spread the virus to numerous other passengers. Based on seating information and laboratory results, investigators were able to map the spread of the virus; this map was interpreted as evidence that the influenza virus was transmitted through the air. More carefully controlled experiments are difficult. No controlled human experiments can be performed for ethical reasons. However, in the 1960s researchers had prisoner 'volunteers' breathe influenza through filters of varying porosity; again, interpretations of results from these experiments were varied, but suggested that influenza could be airborne. Mathematical models of influenza transmission have been constructed, using parameters such as the number of virus particles emitted during a sneeze, the size of sneeze droplets upon emission, the shrinking of droplet size in the air, the distance of transmission of particles of various size, and the number of virus particles likely to reach a 'target' site on recipients. The probability of airborne influenza transmission is considered to be relatively high given reasonable estimates for these parameters.

Even when described at such a coarse grain the various types of evidence regarding the mode of influenza transmission illustrate the problem of discordance. Some scientists argue (using mathematical models and animal experiments) that influenza is transmitted via an airborne route, whereas others argue (based on clinical experience and observational studies) that influenza is transmitted via a contact route. Such discordance demonstrates the poverty of robustness: multiple experimental techniques and reasoning strategies have been used by different scientists, but the results remain inconclusive. A single case does not, of course, demonstrate the ubiquity of discordance; rather, the case is merely meant as an illustration of what is meant by discordance.

If different modes of evidence support contrary conclusions, there is no obvious way to compare or combine such evidence in an orderly or quantifiable way, let alone to compare such a combination of evidence to evidence from a single mode. Philosophers have long wished to quantify the degree of support that evidence provides to a hypothesis. At best, the problem of discordance suggests that robustness is limited to a qualitative notion. And if robustness is a qualitative notion, how should we demarcate robust from non-robust evidence? At worst, the problem of discordance suggests that evidence of different kinds cannot be combined in a coherent way.

One might respond: discordance is not a problem for robustness, since by definition robust evidence is generated when multiple independent modes give the *same* result on the *same* hypothesis. To appeal to discordant evidence as a challenge for robustness simply misses the point. True, but: the problem of discordance is not a knockdown argument against the value of robustness; rather, discordance demonstrates an important constraint on the value of robustness. Robustness, and its corresponding methodological prescription – get more data! (of different kinds) – is obviously valuable. However, this prescription is not something that scientists need to be told – they already follow this common-sense maxim.

That multimodal evidence is often discordant is an empirical claim. Some might think this a weakness of the above argument. However, the opposite is, of course, also an empirical claim – that multimodal evidence is often concordant – and this is an empirical claim which is false. History of science might occasionally provide examples of apparent concordance, but concordance is easier to see in retrospect, with a selective filter for reconstructions of scientific success. Much history of science tends to focus on the peaks of scientific achievement rather than the winding paths in the valleys of scientific effort – at least, the history of science that *philosophers* tend to notice, like Nye's account of Perrin's arguments for atoms, is history of scientific success. Philosophers have focused on the peaks of scientific success, but the lovely paths of truth in the valleys of scientific struggle are often discordant.

Here is a more prosaic way of stating a related worry. Concordant multimodal (robust) evidence for x is sufficient, but not necessary, for a high probability of x. Now, notice two problems that stem from this vague formulation. First, actually specifying the high probability of x depends on principled methods of quantifying concordance and assessing and amalgamating multimodal evidence, which we lack, and thus, we cannot specify the probability of x. That x even has a high probability is merely an intuition. Second, x might be true despite a failure of robustness, but robustness-style arguments do not tell us what to believe in situations of evidential discordance. Franklin suggests that robustness helps resolve discordant data, but I have argued the converse: discordant evidence diminishes the value of robustness. Epistemic guidance is needed most in difficult cases, when multiple independent techniques produce discordant evidence. In such cases robustness is worse than useless, since the fact of multiple modes of evidence is the source of the problem. Real science is often confronted with the problem of discordance.

9.5 Individuating Multimodal Evidence

One advantage of the term "multimodal" is that we can attempt to determine the basis of evidential diversity by determining what modes of evidence are. In other words, clarity on what a mode of evidence is will give clarity on what multimodal evidence is. Understanding what a mode is can partly be determined by knowing what individuates one mode of evidence from another mode. A mode is a type of evidence, of which there can be multiple tokens. For instance, a case-control study is a particular type of epidemiological study design, which can have multiple (infinite) tokens, or instantiations, of the type: two case-control studies identical in all respects except for the number of subjects in each study would not thereby make for two different types of case-control studies, but rather would make for two different tokens of the same type. At first glance, understanding what modes are seems straightforward. Consider the following:

> We have an intuitive grasp on the idea of diversity among experiments. For instance, measuring the melting point of oxygen on a Monday and on a Tuesday would be the same experiment, but would be different from determining the rate at which oxygen and hydrogen react to form water. (Howson and Urbach 1989, p. 84)

While I share this "intuitive grasp" of what multimodal evidence is, it is surprisingly difficult to specify a more clear definition of multimodal evidence. This difficulty is based on the challenge of determining what the proper form of independence should be between modes of evidence. What *form* of independence between techniques – material? theoretical? probabilistic? – is sufficient to individuate evidential modes? What *degree* of independence between techniques – total? partial? – is sufficient to individuate evidential modes? What criteria should we use to individuate modes of evidence? Individuation of modes of evidence is relative to the intended use of the evidence; several uses of multimodal evidence have been suggested in Sections 9.2 and 9.3. Here I consider the independence between modes necessary for robustness arguments.

The individuation problem can be motivated by considering the following simple case, similar to that in the passage from Howson and Urbach. When testing the efficacy of a drug, we might use chemical assays, animal studies, and human trials, each of which we would intuitively describe as a different mode of evidence, and so this would be a case of multimodal evidence. In contrast, performing a particular animal experiment on one day, and then performing the same experiment with all the same parameters again on another day, would not thereby generate two modes of evidence, and so this would not be a case of multimodal evidence (we could call it a case of *mono*modal evidence). Why does the former set of experiments generate *multi*modal evidence and the latter set of experiments only generate *mono*modal evidence? If we had a criterion for the individuation of modes of evidence then we could answer this question, and we would be far along the way to an adequate understanding of what multimodal evidence is and what conditions must be met in order to make a robustness-style argument.

One suggestion is due to Culp (1994): a necessary condition for robustness-style arguments is that modes of evidence should rely on different background theories. It is a commonplace view that evidence is theory-laden, and Culp's suggestion is that the different modes of evidence in a robustness argument must be laden with different theories. But not all evidence is theory-laden in the same way or to the same degree. And sometimes knowing what theory ladens the data is difficult or impossible. Further, I can imagine two pieces of evidence which depend on the same theory for the production of data and interpretation of evidence, and yet which we would call different modes. Consider, for example, all the possible study designs in epidemiology (case-control studies, cohort studies, randomized controlled trials, and so on). Although each of these modes requires particular background assumptions to relate evidence from the mode to a hypothesis, such background assumptions are not necessarily *theories*, if one pedantically reserves this term for high-level scientific abstractions; perhaps some theory is used in interpreting the evidence from these designs, but they are not necessarily *different* theories which laden the evidence from different epidemiological study designs; and yet, these study designs are considered to be different modes of evidence by epidemiologists (though of course they do not use my terminology). Moreover, it is easy to imagine a robustness argument based on evidence from multiple epidemiological studies of different designs. The unit of theory is too coarse-grained to serve as a basis of individuation. Individuation of modes needs a finer-grained criterion than theory independence.

Given that all data is only evidence relative to a hypothesis in conjunction with certain background assumptions, another way to conceptualize the individuation of modes of evidence is by the independence of background assumptions between the modes, relative to a given hypothesis. To individuate two modes, it might be sufficient if the modes share all the same background assumptions except one. One might think that this is not restrictive enough. To consider Tuesday's animal experiment as the same mode as Thursday's animal experiment, besides assuming that the animal experiments followed the same protocol, we must hold several background assumptions on Thursday that we didn't on Tuesday – that the bit of nature under investigation has retained its causal structure since Tuesday, that the different socks which the scientist is wearing on Thursday does not influence the results of the experiment, that the change in the moon's gravity does not influence the results of the experiment, and so on – and yet we would not thereby call these animal experiments two different modes of evidence. Thus it is necessary to have at least a few unshared background assumptions between even tokens of the same mode, let alone between multiple modes.

The other extreme of independence of background assumptions would be when two modes are individuated based on a total exclusivity of background assumptions; that is, when the evidential modes do not share a single background assumption. This might also be too restrictive, since one might think that at bottom all modes of evidence, at least when related to the same hypothesis, must share at least *some* background assumptions. Think of the sensory modalities: vision and touch, though seemingly very distinct modes of sensation, rely on much of the same cognitive apparatus.

Since our knowledge of many background assumptions can be far less than certain, our interpretation of almost any data as evidence for a hypothesis might be an artifactual interpretation based on false background assumptions. A robustness argument based on evidence from different modes, with different background assumptions, might be compelling if the *problematic* assumptions for each mode of evidence – those assumptions which we are uncertain about – were different between modes. Consider a situation in which evidence from a case-control study with high external validity and low internal validity is concordant with evidence from a randomized controlled trial (RCT) with high internal validity and low external validity. To think that both modes of evidence are truth-conducive for a general hypothesis of interest (that is, that both modes of evidence give evidence that is true *and* general, or internally *and* externally valid), it is necessary to hold certain background assumptions for each mode. For the case-control study, a required assumption is that there is no selection bias. For the RCT, a required assumption is that the results are exportable to our population of interest. These evidential modes are individuated rather weakly. They are both human studies at a population level, and as such they share many assumptions, and the statistical analysis of the data from the two modes rely on the same assumptions about population structure. However, the particularly problematic assumptions are the unshared ones. Given that they are unshared, if the two kinds of studies give concordant evidence, that is a reason to think that the unshared background assumptions are not as problematic (in this particular situation) as we would otherwise expect, and so that the evidence is truth-conducive. So *problematic-auxiliary independence* is a good candidate for individuating modes of evidence for arguments based on robustness. The robustness argument for this example would then go as follows. If there was a positive result in the RCT, we might be wrong in assuming that we can generalize its results to a broader population, because of the RCT's low external validity. If there was a positive result in the case-control study, we might be wrong in assuming that the positive result was a true finding, because of the case-control study's low internal validity. But the probability that both studies committed an error is less than the probability that either study committed an error separately.

Thus we can say: it is the background assumptions which we are uncertain about that matter for individuating modes. We can then account for robustness in the following way. A hypothesis is more likely to be true when two or more modes of evidence provide concordant multimodal evidence and the worrisome or problematic auxiliary assumptions for all modes of evidence are independent of each other. At least one problem with attempting to individuate modes based on problematic-auxiliary independence is that we must assume that we can individuate assumptions and determine which assumptions are problematic. This, presumably, can only be done on a case-by-case basis. But how do we know which assumptions are problematic? We could describe the "causal history" or the "mechanism" of a mode of evidence – that is, we could list all the entities and relations involved in the production of the evidence – and then say that if the causal history contains an entity or a relation which is somehow unreliable, then it is the assumptions about that entity or relation which are problematic. This is just pushing the individuation problem back a level: now we have to identify those worrying entities, for which I doubt there is any general criterion of identification. Consider a comparison between electron microscopes and witnesses: evidence from an electron microscope should be construed as being of a different mode than evidence from personal testimony. Two common assumptions thought to be problematic for evidence from personal testimony are based on the witness's capability and the witness's honesty. But a person, the microscope operator, was also involved in the generation of evidence from an electron microscope, and yet we do not normally worry about the capability or the honesty of the microscope operator. It is almost always safe to assume that the microscope operator is honest and capable. Both modes of evidence have, in their causal history, the same type of entity and its associated activity: a person who relays their experience of the world. Despite this similarity, in one mode of evidence the entity has associated problematic assumptions and in the other mode of evidence the entity does not have associated problematic assumptions. Of course, various stories could be told to explain this. My point is that as a criterion of individuation of modes, appealing to problematic background assumptions shifts the burden from specifying a satisfactory and general criterion of individuating modes to specifying a satisfactory and general criterion of identifying problematic background assumptions. This is a burden unlikely to be met.

The prospect of identifying a general definition of multimodal evidence, based on a criterion of individuation between modes, is more difficult than one might have at first thought. This does not entail that, in fact, there are no modes, or that the difference between multimodal evidence and monomodal evidence is illusory or arbitrary. It just means that drawing a sharp demarcation might be impossible. Nor does this mitigate the epistemic importance of multimodal evidence. After all, there does not exist a compelling criterion to individuate sensory modalities, and yet we assume that there are multiple sensory modalities and that having multiple sensory modalities is epistemically important (Keeley 2002). Same with multimodal evidence: we might not be able to come up with a compelling definition of multimodal evidence based on a criterion of individuation for modes, but multimodal evidence remains profoundly important.

9.6 Amalgamating Multimodal Evidence

I suggested that multimodal evidence is said to be important because it is conducive to both certainty, when the evidence from the available modes is concordant, and to uncertainty, when the evidence from available modes is discordant (Section 9.1). But I also suggested that these views of multimodal evidence – that concordant multimodal evidence is conducive to certainty and that discordant multimodal evidence is conducive to uncertainty – are in themselves unsatisfactory. Metaphors like ‘the weight of the evidence’ or ‘robust results’ are usually too vague to warrant assent in the hypothesis in question, and indeed, many scientific controversies are disputes about what the weight of the evidence actually is, or if the results are actually robust or not. If disputants in a scientific controversy had a principled amalgamation function for multimodal evidence, then arguments based on multimodal evidence would be more compelling. Likewise, philosophers making robustness-style arguments would be more convincing if their arguments based on multimodal evidence were supplemented with ways to amalgamate the evidence. Most sciences have crude amalgamation functions for multimodal evidence, but since multimodal evidence is so poorly understood, we have no way to systematically compare or assess the various multimodal evidence amalgamation functions currently in use. I will briefly sketch the contours of what such a function might look like.

To know the impact of multimodal evidence on the confirmation or disconfirmation of a hypothesis, all relevant modes of evidence must be assessed and amalgamated. Modes of evidence should be assessed on several desiderata, including quality, relevance, salience, and concordance. These desiderata have been discussed in detail by others, but to support my argument I will briefly mention them here.[4]

[4] See Galison (1987), Cartwright (2007).

Quality is a straightforward notion which refers to the degree to which a mode is free from systematic errors. *Relevance* refers to the plausibility of the background assumptions that are required to believe that data from a particular mode is evidence for or against a hypothesis. A mode is highly relevant to a hypothesis if data from the mode can be justifiably interpreted as evidence which confirms or disconfirms the hypothesis when such an interpretation requires few implausible auxiliary assumptions. Another important desideratum of evidential assessment is *salience*, which refers to the strength or intensity of results from a mode, or the impact of a unit of evidence on our credence. For example, when testing the efficacy of a new drug to treat depression, if the symptoms in the treatment group improve by five percent compared to the placebo group, that would be a less salient finding than if the symptoms in the treatment group improve by fifty percent compared to the placebo group. Finally, *concordance* is a measure of the degree of consistency of evidence from all the relevant modes for a particular hypothesis. If evidence from all the modes allows for the same inference, given reasonable auxiliary assumptions for each mode, then that multimodal evidence is concordant. Quality, relevance, salience, and concordance do not exhaust the important evidential desiderata, but they are among the most important features of evidence.

Scientists lack systematic methods for assessing quality, relevance, salience, and concordance, though some disciplines have criteria for determining what counts as high quality evidence. For example, the evidence-based medicine movement rank-orders various kinds of studies, with evidence produced by RCTs considered the highest quality of evidence; evidence from prospective cohort studies, case-control studies, observational studies, and case reports normally follow RCTs in descending order of quality.

Different modes of evidence and combinations of modes will satisfy the desiderata to various degrees in different circumstances, by various amalgamation functions. Part of what a good multimodal evidence amalgamation function should do is assess multiple modes of evidence on these multiple evidential criteria: each mode of evidence must be assessed on its quality, relevance, and salience, and the set of the modes of evidence together should be assessed on its concordance. The basis of many scientific controversies can be construed as disputes about differential assessments of these desiderata: one group of scientists might believe that evidence from some techniques is of higher quality or is more relevant to the hypothesis or has greater confirmational salience than other techniques, while another group of scientists might believe that evidence from the latter techniques is of higher quality or is more relevant or salient. For example, Galison argues that one tradition in particle physics considers an image of a "golden event" to be compelling evidence – an observation of a golden event provides strong confirmation to a hypothesis; whereas another tradition in particle physics considers repeatable observations on which statistical analyses can be performed to be compelling evidence.

Abstractly, an amalgamation function for multimodal evidence should do the following: evidence from multiple modes would be fed into the amalgamation function, which would assess evidence on prior criteria (quality of mode), relative criteria (relevance of mode to a given hypothesis) and posterior criteria (salience of evidence from particular modes and concordance/discordance of evidence between modes), and the output would be a constraint on our justified credence. The construction and evaluation of such schemes should be a major task for theoretical scientists and philosophers of science. There currently are functions that combine quantitative evidence from different modes and have a quantitative output, including Demspter-Shafer Theory, Jeffrey conditionalization, and statistical meta-analyses, and there are functions that combine qualitative evidence from different modes and have a qualitative output, including narrative synthesis, meta-ethnography, and there are functions that combine quantitative evidence from different modes but have a qualitative output, such as the evidence hierarchy schemes in evidence-based medicine. An investigation into the methodological virtues and constraints of these functions would be interesting (for example, Stegenga (2011) assesses the purported merits of meta-analysis). With such amalgamation functions, robustness-style arguments might then be more compelling, because there would be a systematic way to guide credence when presented with multimodal evidence. Such functions would be especially valuable when multimodal evidence is discordant. The extent to which robustness-style arguments could be made might be increased if they could be based on multimodal evidence which is not concordant.

9.7 Conclusion

One of the ways that multimodal evidence is said to be valuable is robustness: that is, when multimodal evidence for a hypothesis is concordant, that hypothesis is more likely to be true, or explanatory, or phenomena-saving, or whatever predicate of epistemic success fits most comfortably with one's philosophical inclinations. I have raised several challenges for robustness, the most prominent of which is the ubiquity of discordance. Despite idealizations of scientific success, the world is usually a *rerum concordia discors*. Without the use of compelling schemes to amalgamate discordant multimodal evidence, robustness arguments are vague. Amalgamation functions could provide more constraint on our justified belief in a hypothesis when presented with multimodal evidence.

Appendix: Bayesian Amalgamation

Here I briefly consider how one might amalgamate evidence using a Bayesian approach. Bayesian conditionalization is a rule for revising one's probability of a hypothesis upon receiving evidence. If a scientist learns e, and $p_{\text{old}}(H)$ is the scientist's assessment of the probability of a hypothesis *before* receiving the evidence, then $p_{\text{new}}(H)$ – the scientist's assessment of the probability of the hypothesis *after* receiving the evidence – should equal $p_{\text{old}}(H|e)$. Since this latter term is a conditional probability, it can be calculated using Bayes' Theorem (BT):

$$\text{(BT)} \quad p(H|e) = p(e|H)p(H)/p(e)$$

This suggests a possible way to amalgamate multimodal evidence, based on what is sometimes called ‘strict conditionalization’ (SC): we could update the probability of the hypothesis by sequentially conditionalizing with Bayes’ Theorem for each mode of evidence.[5]

$$\text{(SC)} \quad p_{\text{new}}(H) = p_{\text{old}}(H|e) = p(e|H)p_{\text{old}}(H)/p(e)$$

One could arbitrarily order available modes from 1 to n, and then use Bayes’ Theorem to update the probability of the hypothesis sequentially for each mode, and the posterior probability of the hypothesis after updating on evidence from mode n would become the prior probability of the hypothesis for updating on evidence from mode n+1. The probability of the hypothesis after conditionalizing on the evidence from the first mode would be as above, substituting numerical subscripts for evidence from each mode in place of ‘old’ and ‘new’:

$$p(H|e_1) = p(e_1/H)p(H)/p(e_1)$$

The posterior probability, $p(H|e_1)$, would then be the ‘new’ prior probability, $p(H)$ for updating by evidence from the next mode, e_2:

$$p(H\,|e_2\,) = p(e_2|H\,)p(H|e_1\,)/p(e_2)$$

This sequential updating would continue until the evidence from the final mode, n was used to update the penultimate probability of the hypothesis $p\left(H_{f-1}\right)$ to determine the final probability of the hypothesis $p(H_f)$:

$$p\left(H_f|e_n\right) = p(e_n|H\,)p\left(H_{f-1}\right)/p(e_n)$$

Some Bayesians might consider this approach to be the best way to amalgamate multimodal evidence. Several conditions must be met for this method of sequential conditionalization. For all modes of evidence, all terms in Bayes’ Theorem must be known: that is, for all modes i, $p\,(e_i|H\,)$ must be known; the initial $p(H)$ must be known (this condition has generated much worry, known as the ‘problem of the priors’); and for all modes i, $p(e_i)$ must be known. Determining these terms in practice is often impossible. Consider again the evidence presented in Section 9.4 regarding influenza transmission. What was the probability of observing the pattern of influenza transmission on the landed airplane, conditional on the central competing hypotheses, for example? What was the prior probability of the Contact hypothesis? What was the prior probability of the Airborne hypothesis? Now repeat these questions for the other hypotheses and modes of evidence.

[5] Dutch Book arguments are meant to show that one is rationally required to use SC to learn from evidence.

Also troubling is that Bayes' Theorem requires the scientist using the theorem to know e to be true once e has been observed. In most scientific contexts this is unrealistic. Consider an example given by Skyrms (1986): suppose I see a bird at dusk, and I identify it as a black raven, but because of the evening light, I do not hold the proposition "the bird is a black raven" as my evidence e with perfect confidence (that is, $p(e) \neq 1$). Rather, I might believe e to be true with probability 0.95. Jeffrey (1965) proposed a modification of Bayesian conditionalization to deal with cases in which evidence is uncertain (which, it is reasonable to suppose, is wholly ubiquitous in science). Jeffrey conditionalization (JC), sometimes referred to as 'probability kinematics', is as follows: given multimodal evidence e_i one's updated probability in H, $p_{\text{new}}(H)$, should be:

$$\text{(JC)} \qquad \forall i_{1-n} \sum p_{\text{old}}(H|e_i)\, p_{\text{new}}(e_i)$$

In other words, this is a weighted average of strict conditionalization. To use JC for amalgamating multimodal evidence, one would sequentially update the probability of the hypothesis using JC, similar to the sequential procedure used with SC.

Bayesianism is beset with many well-known problems. This is not the place to rehearse them. But are there any problems with Bayesianism that arise specifically with respect to amalgamating multimodal evidence? A condition of the particular method described above was an arbitrary ordering of the modes. Whatever ordering is chosen should not affect the final probability of the hypothesis. Unfortunately, it is a common complaint against JC that it is 'non-commutative' – the order in which particular pieces of evidence are used to update the probability of the hypothesis makes a difference to the final probability of the hypothesis (see van Fraassen 1989). This problem could be mitigated if there were a way of ordering modes which was superior to others. One might think that if we ordered modes by quality, and used JC on the highest quality mode first and subsequently conditionalized on modes in decreasing order of quality, then the non-commutative property of JC would at least be minimized, because evidence from lower quality modes ought to have a lower impact on the hypothesis anyway. The trouble with this approach is that, despite what some have claimed in particular domains such as evidence-based medicine, there is no general, decontextualized way to order modes of evidence according to a unitary desideratum such as quality. Any ordering of modes will be arbitrary in some important respect. Thus, one cannot resolve the non-commutativity of JC in this way.

Acknowledgments I am grateful for detailed feedback from Léna Soler, Emiliano Trizio and members of the UCSD Philosophy of Science Reading Group, especially Nancy Cartwright. I also benefited from discussion with audiences at the 2008 Canadian Society for the History and Philosophy of Science conference, the 2008 Philosophy of Science Association conference, and the 2008 workshop on robustness hosted by the Archives Henri Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie (Nancy-Université).

References

Achinstein, Peter. 2001. *The Book of Evidence*. New York: Oxford University Press.

Allamel-Raffin, Catherine. 2005. “De l’intersubjectivité à l’interinstrumentalité: L’exemple de la physique des surfaces.” *Philosophia Scientiae* 9(1):3–31.

Bechtel, William. 2006. *Discovering Cell Mechanisms*. Cambridge: Cambridge University Press.

Box, George. 1953. “Non-Normality and Tests on Variances.” *Biometrika* XL:318–35.

Carnap, R. 1950. *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, Nancy. 2007. *Hunting Causes and Using Them*. Cambridge: Cambridge University Press.

Chang, Hasok. 2004. *Inventing Temperature*. New York: Oxford University Press.

Collins, Harry. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.

Culp, Sylvia. 1994. “Defending Robustness: The Bacterial Mesosome as a Test Case.” In *Philosophy of Science Association 1994*, vol. 1, edited by David Hull, Micky Forbes, and Richard M. Burian, 46–57 Chicago.

Douglas, Heather. 2004. “The Irreducible Complexity of Objectivity.” *Synthese* 138(3):453–73.

Fitelson, Branden. 1996. “Wayne, Horwich, and Evidential Diversity.” *Philosophy of Science* 63:652–60.

Franklin, Allan. 2002. *Selectivity and Discord: Two Problems of Experiment*. Pittsburgh, PA: University of Pittsburgh Press.

Franklin, Allan, and Colin Howson. 1984. “Why Do Scientists Prefer to Vary Their Experiments?” *Studies in the History and Philosophy of Science* 15:51–62.

Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.

Hacking, Ian. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.

Hempel, Carl. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

Horwich, Paul. 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.

Howson, Colin, and Peter Urbach. 1989. *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Hudson, Robert G. 1999. “Mesosomes: A Study in the Nature of Experimental Reasoning.” *Philosophy of Science* 66(2):289–309.

Jeffrey, Richard. 1965. *The Logic of Decision*. Chicago: University Of Chicago Press.

Keeley, Brian. 2002. “Making Sense of the Senses: Individuating Modalities in Humans and Other Animals.” *The Journal of Philosophy* 99(1):5–28.

Kosso, Peter. 2006. “Detecting Extrasolar Planets.” *Studies in History and Philosophy of Science* 37:224–36.

Levins, Richard. 1966. “The Strategy of Model Building in Population Biology.” *American Scientist* 54:421–31.

Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Oreskes, Naomi. 1999. *The Rejection of Continental Drift*. New York: Oxford University Press.

Rasmussen, Nicolas. 1993. “Facts, Artifacts, and Mesosomes: Practicing Epistemology with the Electron Microscope.” *Studies in History and Philosophy of Science* 24(2):221–65.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Skyrms, B. 1986. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard University Press.

Staley, Kent. 2004. “Robust Evidence and Secure Evidence Claims.” *Philosophy of Science* 71:467–88.

Stegenga, Jacob. 2011. “Is Meta-Analysis the Platinum Standard of Evidence?” *Studies in History and Philosophy of Biological and Biomedical Sciences* 42:497–507.

Stegenga, Jacob. 2011. "The Chemical Characterization of the Gene: Vicissitudes of Evidential Assessment." *History and Philosophy of the Life Sciences* 33:103–26.

Trout, J.D. 1993. "Robustness and Integrative Survival in Significance Testing: The World's Contribution to Rationality." *British Journal for the Philosophy of Science* 44:1–15.

van Fraassen, Bas. 1989. *Laws and Symmetry*. New York: Oxford University Press.

Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Westman, Robert. 2011. *The Copernican Question: Prognostication, Skepticism, and Celestial Order.* Berkeley, CA: University of California Press.

Wimsatt, William. 1981. "Robustness, Reliability, and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M. Brewer and B. Collins, 124–63. San Francisco, CA: Jossey-Bass.

Chapter 10
Robustness of Results and Robustness of Derivations: The Internal Architecture of a Solid Experimental Proof

Léna Soler

In the Wimsattian definition of robustness as 'invariance under multiple independent derivations' (Wimsatt 1981, reprinted in this book, Chapter 2), the robustness of the invariant result *R* presupposes that the multiple convergent derivations leading to *R* are themselves sufficiently solid (see Chapter 1, Section 1.3). In the present chapter, I address the question of the solidity *of the derivations*.

This will be done through a fragmentary analysis of an experimental derivation involved in a historical episode often described as 'the discovery of weak neutral currents'. This is a well-documented episode, which has notably been studied in detail by Andrew Pickering in his book *Constructing Quarks*,[1] and I will largely rely on the historical material as well as on some philosophical insights offered by this book.

I will proceed along the following road.

To begin, I will introduce the concept of an argumentative line (Section 10.1). The aim is to provide both a general framework in order to characterize the derivations involved in a Wimsattian robustness scheme, as well as the tools needed to specify different kinds of derivations and to distinguish them in the analysis of particular historical cases. Then (Section 10.2) the discovery of the weak neutral current in the 1970s will be reconstructed as involving a robustness scheme composed of three experimental argumentative lines converging on the same conclusion (namely that weak neutral currents indeed are a physical reality). Some reflections will be provided (Section 10.3) with respect to the convincing power of such a scheme and to its widespread realist interpretation.

[1] Pickering (1984). See also Galison (1983). This historical case is very interesting from a philosophical point of view (and has led to controversial interpretations). I discussed it with respect to the issue of a possible incommensurability at the experimental level in Soler (2008c), and with respect to the issue of contingentism in Soler (201X).

L. Soler (✉)
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France
e-mail: l_soler@club-internet.fr

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_10,

Second, I will focus on one of the three experimental argumentative lines involved in the robustness scheme, and I will analyze its internal structure as a 'four floors modular architecture' (Sections 10.4, 10.5, 10.6, 10.7, 10.8, 10.9, and 10.10). After a methodological interlude about the relations between the reconstructed architecture and the level of ongoing scientific practices (Section 10.11), one particular zone of the global architecture, namely the 'Muon noise module', will be examined more closely. I will re-describe it as a prototypical instantiation of the robustness scheme, and will exhibit on the way what I take to be some prototypical features of a convincing robustness scheme (Section 10.12). This will lead to a reflection on the origin of the invariance of the 'something' that is supposed to remain invariant under multiple determinations in a robust configuration. It will be stressed that the 'invariant something' involved in the 'invariance under multiple determinations' formula of robustness, far from being given 'from the beginning', is the result of an act of synthesis characterizable as a more or less creative calibrating re-description (Section 10.13).

Third, the way the elementary scheme of robustness (*N* arrows converging of one and the same result *R*) intervenes inside of the whole architecture of the experimental argumentative line will be analyzed (Section 10.14), and answers will be provided to the initial question of what constitutes the solidity of an argumentative line taken as a whole (Section 10.15). Finally, in the last section, conclusions will be drawn regarding the kind of work the robustness scheme is able to accomplish for the analysis of science, and some potential implications of the chapter with respect to philosophically important issues (such as scientific realism and the contingency of scientific results) will be sketched.

Before starting, one last preliminary remark: The analysis I will propose is not properly speaking an analysis of laboratory practices. It applies, rather, to a level of scientific practices that is emergent with respect to laboratory practices themselves, but that is nevertheless highly important with respect to real scientific developments, and highly relevant with respect to the problem of robustness (see Chapter 1, Section 1.5).

10.1 The Concept of an Argumentative Line

When we ask what makes a given established result *R* solid, we are inclined to appeal to a certain number of supportive elements, that I will name, at the most general level, 'argumentative lines'. This expression remains deliberately vague, since it is intended to encompass any type of derivation, of whatever nature and force, provided that this derivation is believed to support *R*.[2] The term 'argument' seems apt to play this role, since an argument can be either weak or strong, and since the word does not presuppose anything about the kinds of procedures involved.

[2] For the sake of simplification, I only take into account the supportive side of an argumentative line. But actually the concept of an argumentative line, as I conceive it, is intended to be of broader scope, and to encompass, as well, the negative arguments, that is, arguments that play against a result *R*. So at the most general level, an argumentative line is any argument which is relevant to *R*.

Argumentative lines may differ from one another with respect to (at least) three standpoints:

- From the standpoint of their epistemic sphere

  Examples. *Experimental* lines (i.e., supportive arguments in favor of *R* on the basis of performed experiments)/*theoretical* lines (i.e., supportive arguments in favor of *R* on the basis of the content of high-level theories)/*Hybrid* lines (for instance argumentative lines based on simulations).
- From the standpoint of their kind of argument (kind of factors involved, form of the argument. . .)

  Examples. *Analogical* versus *deductive* lines. *Esthetical* lines (i.e., supportive arguments in favor of *R* on the basis of valued esthetic properties such as simplicity, symmetries. . .) contrasted with (what can be comparatively called) *cognitive* lines (i.e., inferences from taken-as-true primitive propositions). . .
- From the standpoint of their force

  At this level we have a rich graduation, reflected in the lexicon familiar to philosophers of science: proof, verification, confirmation, corroboration, and so on.

Obviously, the three characters ‘epistemic sphere’, ‘kind of argument’ and ‘force’ are not independent. In a given historical context, some combinations are especially prototypical and frequently instantiated. For example, today, experimentation is commonly viewed as *the* method *par excellence* in order to establish results in the field of empirical disciplines. In this context, an argumentative line issued from the experimental sphere has great chances to be perceived as a *very strong* argument, that is, as a genuine *proof* – or at least, and more prudently since experimental lines have a high variability in their strength, as a more compelling argument than a purely theoretical line or an argumentative line based on a computer simulation.

Let us now consider from this point of view the particular case of the discovery of weak neutral currents, that is, the case in which the result $R =$ existence of weak neutral currents.

10.2 A Panoramic Analysis of Robustness: The Experimental Argumentative Lines in Favor of the Existence of Weak Neutral Currents

Fortunately, it is not required to know too much about the physics of weak neutral currents to be able to understand the global logic of what I want to develop in this chapter. I will just give some basic elements.

Weak *neutral*-current processes can be defined as weak interactions (scattering or decay) in which *no change of charge* occurs between the initial and final particles. By contrast, a change of charge takes place in weak *charged*-currents processes. At the time of the discovery of weak neutral currents, the situation was currently represented by diagrams of the kind of the ones that appear on Fig. 10.1. Figure 10.1 illustrated the particular case of a neutrino (ν)-nucleon (n) scattering. In the *neutral*-current case, the weak force is mediated by a neutral particle Z_0 and the particles

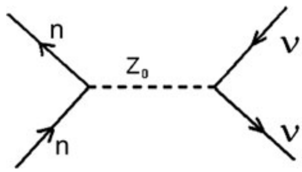

Fig. 10.1 Representation of a neutrino (ν) – nucleon (n) scattering in a Feynman graph

undergo no change of charge. Whereas in the charged-current case, the weak interaction is mediated by a positively charged particle W^+, and the particles undergo a change of charge: the incoming neutrino is transformed into a negative muon μ^- at the upper vertex, and the neutron is changed into a proton at the lower one.

The period commonly associated to the discovery of weak neutral currents is, roughly, between 1972 and 1975.

Which argumentative lines will be marshaled, from today's standpoint, if one asks what gives robustness to the proposition, currently viewed as well-established, that weak neutral currents exist?

If we simplify the situation, that is, if we only consider the *experimental* argumentative lines, and if we only consider the experimental argumentative lines which played *in favor* of weak neutral currents *during the short period of 1972–1974*, we will invoke (at least) three favorable lines:

- The experimental line I will call 'Gargamelle' (L_1)
  This line investigates the weak neutral current reaction 'ν + nucleon $\rightarrow$ ν + hadrons' (where ν is a neutrino[3]) with a 'visual' detector (in that case a giant bubble chamber named 'Gargamelle') (Hasert 1973b, 1974).
- The experimental line I will call 'NAL' (L_2)
  This line investigates the same weak neutral current reaction 'ν + nucleon $\rightarrow$ ν + hadrons' with an 'electronic' detector (at the National Accelerator Laboratory) (Benvenutti 1974).
- The experimental line I will call 'Aachen' (L_3)
  This line involves another weak neutral current reaction 'ν + electron $\rightarrow$ ν + electron' and a bubble chamber (Hasert 1973a).

These three experiments show undeniable differences from one another. Indeed they at least differ two by two: on the one side with regard to the experimental design (visual detectors for Gargamelle and Aachen, electronic detector for NAL); and on the other side with regard to the type of weak neutral current reaction.

Here, we seem to get a perfect exemplification of a robustness scheme à la Wimsatt, in the case in which the multiple derivations correspond to *experimental* argumentative lines. One seems thus entitled to represent the situation by means of the diagram of Fig. 10.2.

[3] For more details about this interaction see below Section 10.5.

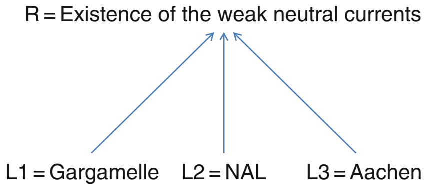

Fig. 10.2 The retrospective panoramic scheme in the case of robustness of the weak neutral current discovery

10.3 Revealing the Tacit 'Argument' That Gives the Robustness Scheme Its Convincing Power and Places It in a Position to Play the Role of a Springboard Toward Realist Attributions

What is it that provides this retrospective panoramic scheme with its convincing power? What is it that so strongly pushes us (philosophers of science as well as actual practitioners of science) to feel that in such a configuration, the result R is indeed robust in an intuitive, 'pre-Wimsattian' sense? What is it that works as a 'justification' of the robustness of the result R in such a situation?

The same kind of reasons, I think, lies behind the current intuitive attributions of robustness by scientists and philosophers of science on the one hand, and on the other hand the Wimsatt's decision to *define*, for epistemological purposes, a precise, explicit sense of robustness, namely the invariance of R under multiple independent derivations. It seems to me that these reasons are implicitly related to an argument of the 'no-miracle argument' kind.

In its most naïve and most convincing form, this argument conceives the derivations as independent, both logically-semantically with respect to their content, and historically with respect to their empirical implementation.[4] On the one side, people develop L_1... And find R. In parallel, people develop L_2, very different in content from L_1... And find R. In parallel, people develop L_3, very different in content from L_1 and L_2... And find R. Then, the results are confronted: three times R! (see Fig. 10.3). Three experiments wildly different in content, each leading to one and the same result, three different and independent argumentative lines converging on one and the same result R... (it is at this point that the Fig. 10.3 is converted into Fig. 10.2). This would be an extraordinary coincidence – a 'miracle' – if it was by chance... It is much more plausible to conclude that R is indeed, 'in itself' or 'intrinsically' robust...

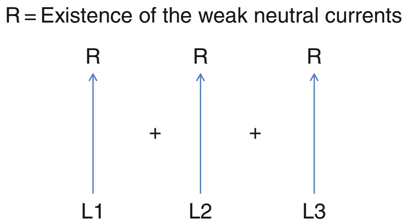

Fig. 10.3 A common reading of the retrospective panoramic scheme of robustness

[4] See Chapter 1, Section 1.8 for more developments on this distinction between two kinds of independence.

At this point, if we ask what, exactly, does 'intrinsically' mean in this context, we are quasi-inevitably led to realist intuitions. To say that three times R by independent lines cannot be by chance, is to mean that something outside us must have participated, conspired to precipitate R... This intuition that we have bumped into something outside us can be expressed through different, more or less strong formulations: R has been obtained *because* R is indeed a *genuine* characteristic of the object under study (in contrast to an experimental artifact, a mistake of the subject of knowledge...)... Or stronger: R has been obtained because R has been *imposed* by the object under scrutiny... R is *objective*, if not (at least approximately) *true*...

Of course, *from a philosophical point of view*, the leap from the claim that we have several, indeed sufficiently independent derivations leading to one and the same R, to the objectivity if not the truth of R, would need to be argued (and is indeed highly questionable[5]). But here, I just want to point to the kind of intuition that lurks behind the robustness scheme and is, I think, the common source of its force and convincing power.

10.4 From the Panoramic Scheme of Robustness to the Internal Structure of One of Its Derivational Ingredient

One could, relying on the example of the discovery of weak neutral currents, discuss the validity of the robustness scheme at the panoramic scale. But despite the great interest of this example in this respect (see my remarks in Chapter 1, Section 1.8.3), that is not what I want to do in this chapter. What I want to do is, from the panoramic scheme, to zoom in, and to focus on *one* of the argumentative lines, namely, the Gargamelle line.

The panoramic scheme offers a view of the situation *at a given scale*. At this scale, it adopts a simplified representation, one that consists in identifying each of the experimental lines with a monolithic unity: *one* experimental argument, *this* experimental proof (represented on the scheme by a *single* arrow).

But of course, looking more closely, zooming on a supportive arrow, we discover a complex internal argumentative structure.

My aim, in this chapter, is to open the procedural black box represented by the Gargamelle arrow and to analyze its internal structure, with the intention of drawing some general conclusions about the solidity of procedures and the solidity of results.

[5] One reason to think that the leap is questionable is the multiplicity of the historical cases in which schemes of robustness indeed obtained at a given research stage S_1 have been dissolved at a subsequent stage S_2. In that movement, the node R first taken to be objective/true on the basis on the robustness scheme indeed available at S_1, is, subsequently at S_2, re-described as an artifact, a human mistake or the like. For insights about this configuration and a historical example, see the paper of Stegenga, Chapter 9 (as he rightly concludes: "concordant multimodal evidence can support an incorrect conclusion". For other kinds of possible arguments against the jump from robustness to realism, see the concluding section of this chapter, and Chapter 1, Section 1.8.3.

The following reflections will mainly rely, in terms of primary sources, on an article published in 1974, which is, along with a few others, commonly considered to announce the discovery of weak neutral currents.[6]

10.5 Some Preliminary Technical Elements in Order to Understand the Gargamelle Line: The Charged Currents as an Ally in Order to Convert the Machinic Outputs into Theoretically Defined Neutral Currents Interactions

First, let me introduce a few preliminary elements required in order to understand the Gargamelle argumentative line.

In the Gargamelle experiments, the (hypothetical) weak neutral reaction under study is:

$$\nu + \text{nucleon} \rightarrow \nu + \text{hadrons} \quad \text{(NC)}$$

A neutrino-nucleon scattering produces, as secondaries, a neutrino and a shower of hadrons. I will call this particular kind of weak neutral current 'NC'.[7]

How can the NC reaction be experimentally identified?

In the visual experiments of the Gargamelle kind, the data resulting from the experiments are, at the rawer, less-interpreted level, photographic images on a film. I will call these pictures the 'instrumental outputs' or 'machinic ends' (playing on the double sense of 'end' as the output and the aim), and I will equate them to the 'zero degree' of experimental data (what is sometimes called "raw data" or "marks", see for example (Hacking 1992)).

Now, the correlation between visible tracks on the film on the one hand, and physical, theoretically defined events on the other hand, is not always an easy matter. Neutral particles, in particular, leave no visible tracks on the film. Hence the experimenters cannot but *infer* the presence of this or that neutral particle *from* the visible tracks of charged particles with which these neutral particles have interacted.

Since the hypothetical neutral currents NC involve *two* neutral particles, the incident neutrino and the outgoing neutrino, the experimental identification of an NC

---

[6] Hasert (1974). See also Hasert (1973b) (which is roughly the same paper but with fewer details), (Benvenutti 1974) (which target the same weak neutral interaction through an *electronic* experiment – what I have called above the NAL line), and (Hasert 1973a) (which is about another kind of weak neutral current, the muon-neutrino *electron* scattering, investigated with a bubble chamber – what I have called above the Aachen line). I do not claim, of course, that these papers exhaust the publications commonly considered to announce the discovery of weak neutral currents.

[7] A neutrino is a lepton, that is, a particle subject to weak interaction, whereas hadrons are particles subject to the strong interaction. To be more precise, the incident particles involved in the interaction are muonic neutrinos. The reaction corresponds to a neutral current, since no change of charge happens: both the incoming and the outgoing leptons are neutral particles.

is a delicate task. A major concern of experimenters in the 1970s was the possibility of mistaking a pseudo-NC event *produced by a high-energy neutron* for an authentic NC event *produced by neutrinos*. This problem was known as the neutron-background problem.

In this context, a crucial point that I want to stress, is that the experimental identification of the NC-events involves, *in a decisive way*, *events other* than the NCs. Indeed, an entire set of other events, that I will call 'the space of other relevant events'.

These other events can play the role either of allies (in the sense that they are helpful with respect to the identification of the NC events) or of parasites (in the sense that they could be confused with a NC event, that they work as background noise). In this chapter I will only be able to introduce a very small number of the events pertaining to the space of other relevant events. But there is one of them, absolutely constitutive of the Gargamelle line, which, because of its role of primary ally, has to be mentioned.

This is the *charged* current reaction symmetrical to the neutral reaction NC under study:

$$\nu + \text{nucleon} \rightarrow \mu^- + \text{hadrons} \quad \text{(CC)}$$

Where μ^- is a negative muon. I will call this reaction 'CC'.

Compared with a NC process, the same incoming particles are involved; but here we find, in the outgoing particles, in addition to the hadrons, a negative lepton instead of a neutrino.

How does the CC-interaction play its role of ally with respect to the NC interaction? At the time, the CC were, contrary to the NC reactions, assumed to exist, and the CC-interactions were much better known, both theoretically and experimentally, than the NC interactions. These circumstances led the experimenters to transform the initial problem 'Experimental detection of the NC reaction', into this other problem: 'Experimental evaluation *of the ratio NC/CC*'. This methodological strategy – resorting to a ratio, one of whose terms is better known – exemplifies a paradigmatic strategy, and gives the CC interaction the status of a primary ally.

10.6 The Global Architecture of the Gargamelle Argumentative Module

Let us now turn to the analysis of the internal structure of the Gargamelle experimental line.

10.6.1 From the Line to the Module

When we analyse the constitution of any unitary argumentative line, we are, intuitively, inclined to replace the image of the arrow that was natural in a panoramic overview, with the image of the box or the module.

I will thus consider the Gargamelle line as a module, and will represent its internal architecture by a series of sub-modules included one in the others, like Russian dolls (but with more complex combinations of inclusions).

Before applying this representation, some brief remarks must be made about the methodological principles that govern the *individuation* of a module *as a* modular *unit at a given scale*. A module will be individuated and defined as a unit on the basis of its aim: on the basis of the question it is intended to answer, of the problem it tries to solve.

In that vein, the argumentative line 'Gargamelle' can be instituted as a modular unit of the same name (i.e. the Gargamelle module) defined by the aim-question: 'is the NC-interaction experimentally detected with the Gargamelle bubble chamber?'

Similarly, what I have called, in Section 10.2, the "panoramic retrospective scheme of the robustness of R", with $R =$ existence of weak neutral currents, can be viewed, by zooming back from the Gargamelle module and by considering the situation at a broader ("panoramic") scale, as a unitary module individuated by the aim-question: 'Are weak neutral currents experimentally detected?'

As any situation can always be analyzed in different manners in terms of the structure of the aims, it must be stressed from the outset that the modular architecture that will be proposed in what follows is not univocally and inevitably imposed by the objective text of the 1974 paper. Sometimes – when turning to certain parts of the Gargamelle argument – the conviction forced itself upon the analyst, intuitively, that it has to be *this* unique decomposition, that it cannot be anything else… But when turning to some other parts, several options appear possible, and hesitations arise concerning the most adequate or relevant one. In any case, the analyst always has a certain degree of freedom, and the structural configuration finally adopted always depended on him for certain decisions.

10.6.2 The Gargamelle Module as a Four Floors Building

Here is a sketch of the global architecture of the Gargamelle line as I have decided to analyze it.[8]

[8] In what follows, the "Gargamelle line" refers to the road that goes from the machinic outputs (the photographic images on the film) to the conclusions about 'what they say' in terms of the NC/CC rate. What happened before in order for the experimenters to be in a position to obtain and trust the corresponding pictures (the history of the construction of the bubble chamber, the history of the knowledge that was required in order to conceive something like a bubble chamber and so on…), is not taken into account. In other words, the Gargamelle line as it is analyzed in what follows is restricted to what is often called "data analysis". Obviously, the Gargamelle experimental derivation could be understood in a broader sense, including elements of the anterior history of science that are presupposed in order to take what I called the 'degree zero' of the experimental data (what appears on the film) as reliable data. In relation to the delimiting choice made in the present chapter and to the adopted re-description of the Gargamelle line so delimitated as a four-floors edifice, we could say that this four-floors building is not suspended in the void but rests on a deep and structured underground.

At a first level of analysis, the Gargamelle argumentative module can be split into four big boxes or sub-modules. These sub-modules can be seen as four logical moments or logical steps,[9] defined by their intermediary specific aims. In a first approach, these four steps will be considered as logically successive, and will be ordered sequentially from the bottom to the top along a vertical axis.

This bottom-up visual representation is intended to suggest graphically that what is logically prior (lower in the diagram) largely conditions, not to say is irreducibly constitutive, of what is logically posterior (higher in the diagram).

I will now present, in an unavoidably concise way, each of the four floors of the whole architecture.

The architecture will involve the four following floors (see Fig. 10.4):

- As the first floor: a box 'Selection of the basic data for the analysis' (or for short, 'Selections')
- As the second floor: a box 'Coarse estimations of the number of events within the selected data' (for short, 'Coarse estimations')
- As the third floor: a box 'Refined estimations after correction of the coarse first estimations obtained' (for short, 'Refined estimations' or 'Noise')
- Finally, as the fourth floor: a box 'Confrontation between the experimental results obtained and the predictions of the high-level theories' (for short, 'Confrontation with high-level theories').

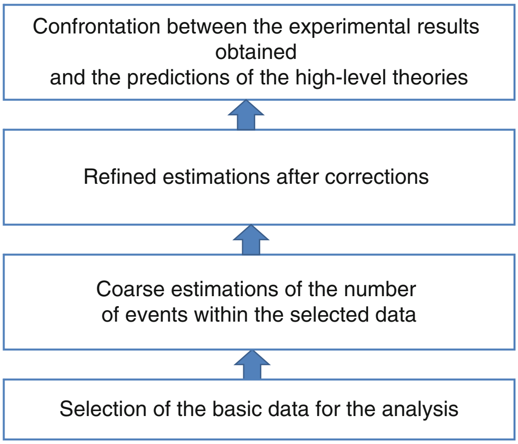

Fig. 10.4 The Gargamelle module as a four floors building

[9] For more considerations about the epistemological status of these logical steps, especially with respect to their relation to the chronology of actual scientific practices, see below Section 10.11.

10.7 The First Floor: The Module 'Selections'

10.7.1 The Internal Constitution of the Ground Floor: Four Parallel Selections

At the end of the Gargamelle experiments, we have 290000 photographic images recorded with the giant bubble chamber Gargamelle. But all the tracks visible on the photographic film are not retained. Only a sub-part of the totality of the machinic ends are taken into account.

Four operations of selection or filtering are performed. I will represent them by four parallel sub-modules (see Fig. 10.5), without having space to describe their content. Just to give an example: in the module 'Energetic cut at 1 GeV', from the start experimenters get rid of all the events whose total energy is below the threshold of 1 GeV.

The aim of such selections is almost always to exclude at once, from the entire set of the machinic outputs, a sub-set of tracks that are judged too ambiguous, either because they are not clearly readable in terms of individual geometric properties, or because their population is deemed infected by a huge number of pseudo-events. In other words, the aim is to extract a set of tracks whose interpretation is globally more reliable, in such a way that it becomes less likely to make mistakes in the counting of the potential tracks of NC.

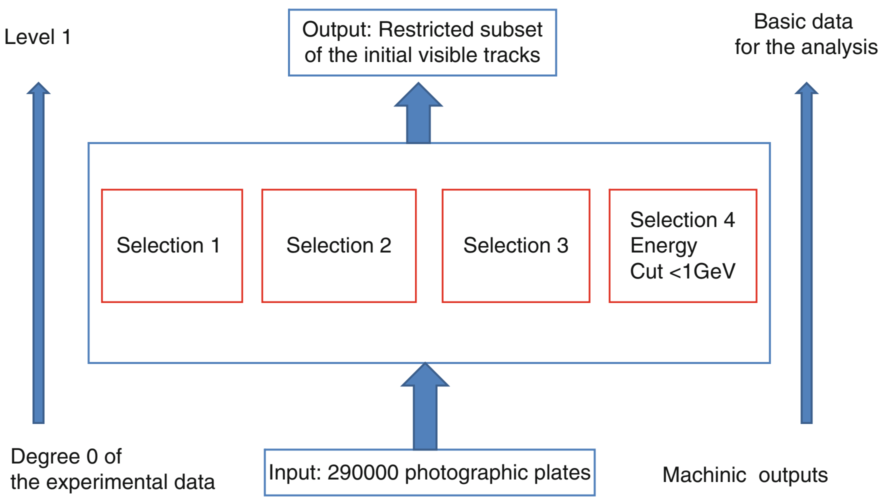

Fig. 10.5 The first floor: The module 'Selections'

But if this is the aim, experimentalists are never sure that it is indeed achieved. The filtering operations aim at eliminating some confusions, but they can themselves be sources of mistakes. For example, if the energy cut at 1 GeV is too severe, real NC-events might be artificially eliminated, and the risk is to conclude mistakenly that weak neutral currents do not exist. But if the cut is too permissive, too many pseudos might be taken for authentic NC-events, and the risk is, this time, to conclude mistakenly that weak neutral currents do exist.

The four preliminary selections performed appear variably problematic to the experimenters' eyes. But, whether problematic or obvious, the operations involved in each of the four sub-modules at the first floor have essential repercussions on the conclusions that will be drawn at upper floors. They are completely constitutive of the final answer that will be given to the question of the detectability of the NC reaction.

10.7.2 The Resultant Assessment of the Four Cutoff Operations: At the Exit of the 'Selections' Module

As the input of the 'Selections' module, we have the totality of the machinic outputs, namely all the visible tracks that appear on the 290000 photographic images (see Fig. 10.5). At the output of the 'Selections' module, after concatenation of the four specific selections, we have a restricted sub-set of all the visible tracks.

I will describe the result of these four constitutive operations of global filtering as the institution of a new layer, that I will call 'level 1 of the experimental data' (of course the number one acquires its sense only relatively to the level zero). Here the level 1 can be specified as the 'level of the basic data for the analysis', for it is at this level that the experimenters are going to evaluate the number of track patterns that could be manifestations of NC events. (See Fig. 10.5 for a schematic overview of the first floor).

The basic data for the analysis, that is, the output of the big 'Selections' module viewed as the first floor of the construction, will constitute the input of the module situated just above in my ascendant vertical representation.

10.8 The Second Floor: The Module 'Coarse Estimations'

I identify this module, viewed as the second floor of the edifice, as a modular unit by the aim: to build, from the pre-selected sample of visible tracks associated with level one, a first coarse estimation of the relevant events, namely, primarily the NC and CC events.

The second floor will be portrayed as a duplex (see Fig. 10.6).

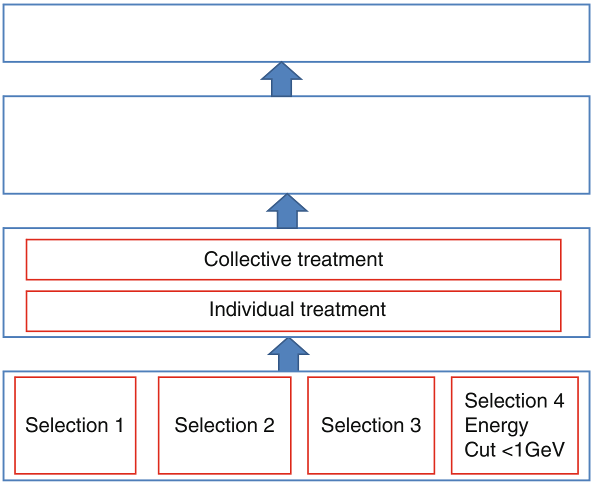

Fig. 10.6 The second floor as a duplex

10.8.1 The Lower Part of the Duplex: The Module 'Individual Treatment'

The lower part of the duplex, that I will call 'Individual treatment', will be defined as a unitary sub-module by the aim: To count, on the photographs, the individual track-events of each type.

In order to achieve this goal, the experimenters specify, for each type of relevant theoretical event (NC, CC. . .), the observable characteristics a pattern of tracks must necessarily satisfy to be classified, at least provisionally as a first plausible hypothesis, as a NC-event, or a CC-event, etc.

Next the experimenters count, within the pre-selected sample of the machinic ends, the number of track-patterns that satisfy the criteria of experimental identification defining what I will call a NC-*candidate* and a CC-*candidate*.

As a result, they find:

NC-Candidates = 102
CC-Candidates = 428

At this stage, as the experimenters stress, "The number of NC events is large".

10.8.2 The Upper Part of the Duplex: The Module 'Collective Treatment'

The upper part of the duplex 'Coarse estimation' will be called 'Collective treatment', and will be defined as a unitary sub-module by the aim: To check, through the examination of *collective* properties of the populations of events, that no major mistake has been made in the previous step corresponding to the *individual* experimental identification of the NC-candidates.

In order to achieve this aim, the general strategy of the experimenters is to institute two privileged points of comparison, to which the 102 NC-candidates are confronted: a positive point of reference, namely the CC-candidates; and a negative point of reference, namely the neutral hadrons (and on the front line the neutrons).

The logic of the argument can be reconstructed as follows: If most of the 102 NC-candidates identified on the film are *authentic* NC-events produced by neutrinos, it is expected that their collective characteristics will present, *statistically*, some *essential similarities* with the collective characteristics of the 428 CC-candidates. Whereas sharp differences (of a partly determined type) are expected if a majority of the 102 NC-candidates actually are *pseudo*-NCs induced by neutral *hadrons* and not by neutrinos.

In this module, one can see precisely how the CC interaction plays, *in concreto*, its role of ally *as an experimental standard*. The collective properties of the population of 428 CC-candidates identified on the film are turned into an experimental norm. They show what collective properties a population of authentic NCs *should* have. They work as benchmarks.

The aim of the 'Collective treatment' sub-module is achieved by applying this general strategy to four different collective characteristics: the spatial distribution, the energetic distribution, the angular distribution, and the mean free path of interaction. The four corresponding investigations will be represented as four parallel sub-sub-modules (see Fig. 10.7).

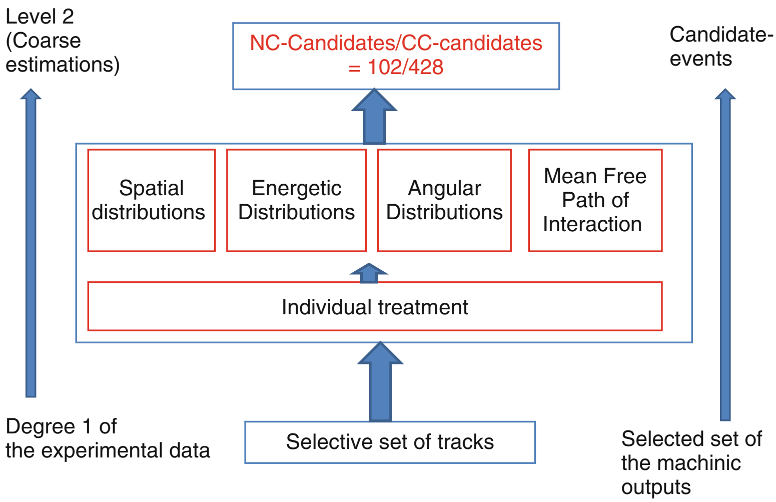

Fig. 10.7 The second floor: The 'Coarse estimations' module

10.8.3 The Resultant Assessment of the Upper Part of the Duplex: At the Exit of the 'Coarse Estimations' Module

Each of the four statistical tests corresponding to the four parallel sub-sub-modules shows a global resemblance between the collective behavior of the NC-candidates and the collective behavior of the CC-candidates with respect to the physical variable involved. Thus, all four sub-modules of the 'Collective treatment' module are favorable to the hypothesis that most of the 102 NC-candidates indeed are *authentic* NC-events rather than pseudo-NC.

Thus at the exit the module 'Coarse estimations' (see Fig. 10.7), the intermediary conclusion is the following first raw evaluation of the NC/CC rate:

$$\text{NC-Candidates/CC-candidates} = 102/428$$

I will describe the level of this output as a 'level 2 of the experimental data'. In this example it corresponds to the level of the coarse estimations. The typical epistemic status of the conclusions reached at level 2 is '*candidates'*, a term which aims to indicate the *still approximate and provisional character* of the rate value retained. (See Fig. 10.7 for a schematic overview of the second floor).

The rate '102/428' of the NC/CC-candidates that holds as the output of the first floor 'Coarse estimations', in itself "large" as stressed by the experimenters, will constitute the input of the upper floor, the second floor.

10.9 The Third Floor: The Module 'Refined Estimations' (Or 'Noise')

10.9.1 Identifying the Background Noise

The module corresponding to this second floor is defined by the aim: To obtain a refined, realistic estimation of the NC/CC ratio.

In order to achieve this goal, the coarse estimations obtained at level 2 must be corrected, taking into account the influence of new elements of the space of the relevant events. This time, these elements play the role of parasites, and in the 1974 paper, they are altogether categorized as "background noise".

Four different sources of noise are mentioned and respectively treated. The module 'Refined estimations' or 'Noise' can thus be decomposed in four parallel sub-modules that I will only have the space to mention: 'Noise of low momentum muons'; 'Noise of cosmic rays'; 'Noise of neutral hadrons from the primary beam'; and 'Noise of neutral hadrons from the secondary beam' (see Fig. 10.8).

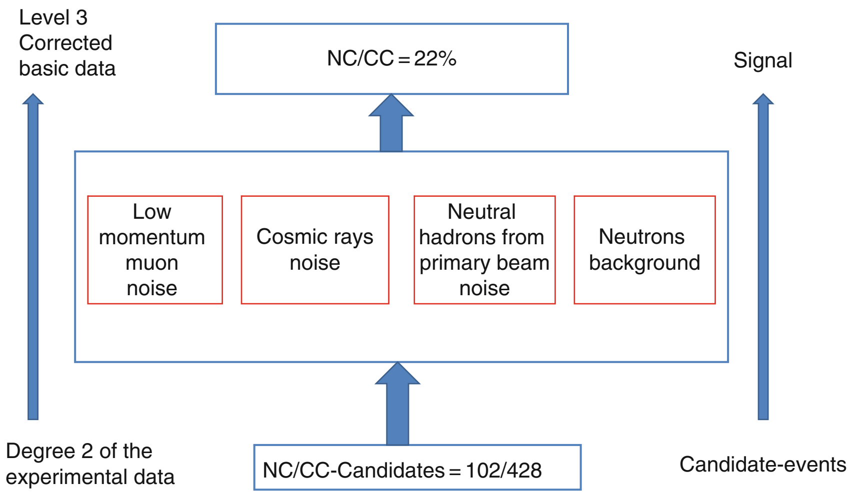

Fig. 10.8 The third floor: The 'Refined estimations' module

As the output of each sub-module 'Noise', a numeric estimation of the number of pseudo-events of the relevant type is obtained.

10.9.2 The Resultant Assessment of the Four Sub-modules 'Noise': At the Exit of the 'Refined Estimations' Module

The global result of the four parallel sub-modules 'Noises' is obtained by conjunction: simplifying, one subtracts, from the number of the 102 NC-candidates obtained as the output of the second floor, the different numbers obtained at the third floor for each type of pseudo-events.

At the exit of the module 'Refined estimations', the conclusion is, finally (see Fig. 10.8):

$$NC/CC = 22\%.$$

I will describe this whole stage of identification and processing of the different sources of noise as the constitution of a new level, the level 3 of the experimental data, that I will call the level of the *corrected basic data*.

On the level of the corrected data, the conclusions are viewed as the 'best estimations that can be achieved in the current state of the research'. The 'something' that is quantitatively evaluated at level 3 has another epistemic status than the one of 'candidate' which was typical of level 2. It goes with a stronger realistic pretension: no more 'simply' a candidate... But a '*real*' something... At the level of the *corrected* experimental data, experimenters claim to have identified an *authentic* phenomenon (as opposed to an artifact), even if what is at stake remains in many respects hypothetical and only partially characterized from the theoretical point of view.

In order to grasp the status of what is at stake, I will use the category of the 'signal'. (See Fig. 10.8 for a schematic overview of the third floor).

As I conceive it, the category of the signal is intended to name the highest level of the experimental data ('data' already highly elaborated and interpreted as one can see). Whatever its number, the signal corresponds to the *surface* of the stratified experimental analysis, to the final point of the argument as an *experimental* argument. That is why in the 1974 paper, the conclusions associated with my category of signal are presented in a rubric entitled "results".

Beyond the stratum of the signal, we leave 'what experiment says', to enter into the sphere of the theoretical interpretation of what has been experimentally extracted. True, to talk about a *NC*-signal is already to project a given theoretical interpretation of the 'something' that has been experimentally extracted. But at the level of the signal, this interpretation remains presumptive. So, on the whole, what the expression 'S-signal' exactly picks out, is: 'the something that has been extracted from the instrumental outputs and partially characterized, and that can be *potentially* interpreted *as an S*'.

10.10 The Fourth Floor: The Module 'Confrontation of the Experimental Signal with High-Level Theories'

I don't have the space to describe the fourth floor. The aim of the corresponding module is to confront the NC/CC signal of 22% to what high-level theories say – and especially one of them, the Weinberg-Salam theory, which assumes the existence of the NCs.

At this floor the experimenters begin to stress that "The neutral current hypothesis is not the only interpretation of the observed events". Then they list and investigate very partially some possible interpretations (these interpretations being re-describable as four parallel sub-modules, see the upper floor of the Fig. 10.9). Finally they arrive at this conclusion, striking for its cautiousness: "Interpreting these events as induced by neutral currents", they appear "compatible" with the Weinberg-Salam theory of weak interactions.

This very cautious formulation, which, clearly, is far less than an outright declaration of existence proper, fits well with the Wimsattian scheme of robustness. Indeed, the conclusions of the 1974 paper are only supported by *one single* type of experiment, whereas the Wimsattian scheme of robustness requires *several different* types of experimental lines (or in the terminology of Catherine Allamel-Raffin: requires *inter-instrumentality* (Allamel-Raffin 2005). It is then perfectly congruent with the scheme, and even *required* according to it, that the conclusion

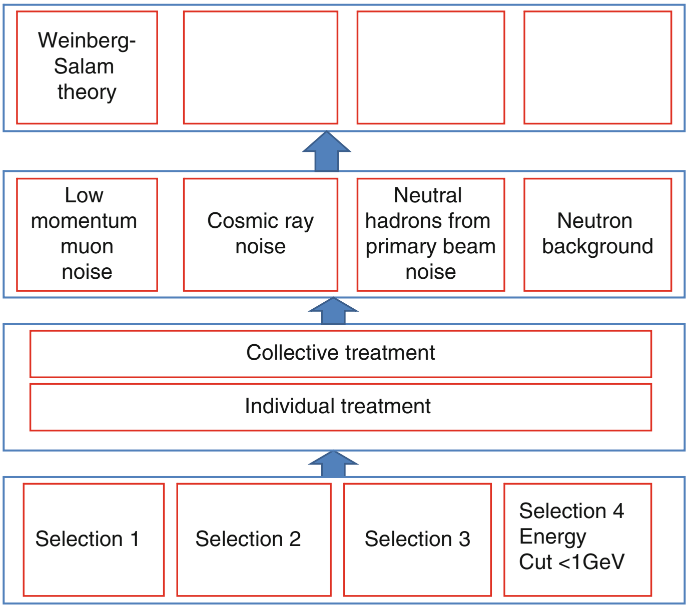

Fig. 10.9 Decomposition of the Gargamelle module in sub-modules

of the chapter remains fragile, only plausible, associated with a weak degree of robustness.

10.11 Methodological Interlude: Some Remarks About the Relation Between the Gargamelle Architecture and Scientific Practices

Before continuing the analysis of the Gargamelle line, I would like to say a few words about the gap between the level of the previous analyses and the level of science in action.

The previous analyses, and the architectural representations that go with it, are based on the text of a published article. They are, thus, verbal re-descriptions and visual re-presentations of a reality that is akin to the level of emergent stabilizations. As I argued in Chapter 1, Section 1.5, the study of what holds at this emergent level is relevant to the problem of robustness.

But it is important to sharply separate the two levels, and to *refrain from equating the logics associated with each of them*. Indeed, the succession of the four big modules along the vertical axis *cannot be equated* with four successive steps that, chronologically, would have been taken *as such consecutively* by practitioners. It cannot even be equated with four *logical* moments that would have been *thought* *and performed as such in this order* by real practitioners. The four floors architecture is an emergent, highly simplified and largely reordered one with respect to the multiple constructions built all along the actual path. And the simple, sequential process exemplified by the different floors is not a good model of practices. In real practices, what is involved is a reticular logic with retroaction loops and multiple restructuring along the path.

Having sketched a (largely simplified) overview of the internal architecture of the Gargamelle emergent argumentative module (see Fig. 10.9 for a graphical overview), I will now zoom in again, and have a closer look at the internal structure of some of the sub-modules constitutive of the edifice.

10.12 The Internal Structure of the 'Muon Noise' Module: A Prototypical Example of the Robustness Scheme

I will begin with the sub-module 'Noise of muons with low momentum', since intuitively, one is inclined to see it as a perfect illustration of the robustness scheme.

The noise here identified, that is, the feared risk of mistake, is the following. A track of a negative muon of low momentum (<100 MeV) could be confused with a track of a short stopping proton (that is to say a hadron). Now, the presence or the absence of a muon within the outgoing hadrons is precisely what distinguishes the CC and NC interactions. If a low momentum muon is taken as a hadron, one will count as a NC-candidate what is actually a CC (a pseudo of the type: 'CC with a low momentum muon'). We see here how an ally can switch into an enemy (here a parasite).

On the ground of this analysis, the problem to solve, which defines the 'Muon noise' module as a modular unity, can be formulated as follows: Estimation of the number of pseudos of the type 'CC with a low momentum muon' within the experimental sample of the 102 NC-candidates.

"The magnitude of the effect may be estimated (. . .)", write the experimenters in the 1974 paper. How? The estimation involves three distinct parallel sub-modules (see Fig. 10.10).

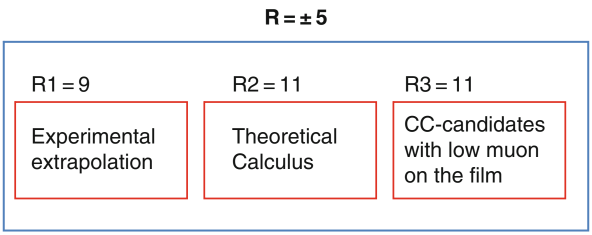

Fig. 10.10 The 'Muons noise' module, as a prototypical exemplification of the robustness scheme

(A) In the first, which I will label ‘Experimental extrapolation’, experimenters extrapolate “the observed muon spectrum to zero energy”. This spectrum is constituted by muon-candidates of energy superior to 100 MeV, thus of tracks that are *not* subject to the muon-proton ambiguity.

Upshot and output of the sub-module ‘Experimental extrapolation’: “This procedure predicts a misclassification of 9 events”.

Conclusion: $R_1 = 9$ (see Fig. 10.10).

(B) In the second parallel sub-module, which I will label ‘Theoretical calculus’, the number of low-momentum muons are determined by a theoretical calculus. The calculation is not detailed in the chapter, but some theoretical hypotheses on which it is based are explicitly mentioned (“a theoretical calculation assuming scaling and correcting for non-zero muon mass”).

Upshot and output of the sub-module ‘theoretical calculus’: 11 events.

Conclusion: $R_2 = 11$ (see Fig. 10.10).

(C) In the third parallel sub-module, experimenters examine, on the film, how many events already classified as CC-candidates have, in their secondaries, a muon-candidate with a momentum inferior to 100 MeV.

Upshot and output of the argumentative sub-module ‘CC-candidates with low-moment on the film’: 11 events.

Conclusion: $R_3 = 11$ (see Fig. 10.10).

(D) The resultant assessment of the three sub-modules: at the exit of the module ‘Muon noise’

“The correction to be applied is 0 ± 5 events”, write the experimenters. They found 9, 11 and 11. They retained the value ‘10’.[10] That is, they conclude that the mistake in question concerns 10 events. Since they don’t see any reason that would favor the erroneous overestimation of the NC-candidates at the expenses of the CC-candidates, or the opposite mistake, they distribute symmetrically the amplitude of ‘10’ between plus and minus.

Upshot and output of the module ‘Muon noise’: the experimenters could have over- or under-estimated by 5 events the initial count of the 102 NC-candidates.

$R = \pm 5$ (see Fig. 10.10).

The procedure corresponding to the ‘Muon noise’ module exemplifies characteristic traits of the Wimsattian scheme of robustness, at least if we accept the following re-description of its content:

- First, we find several parallel derivations (namely three) for the estimation of one and the same magnitude (the low momentum muon background).
- Second, each of the three parallel approaches seems to be taken as ‘in itself sufficiently reliable’ (this is of course implicit in the paper but can be

[10] The experimenters don’t give any further explanation, but I will consider in more detail below (Section 10.13) the content of the process and the motivation that might have led them to retain this value ‘10’ from the three values 9, 11 and 11.

assumed on the basis of the absence of any discussion devoted to the matter, joined to the extreme brevity with which the whole issue is settled).

- Third, the three derivations involve notable differences. I cannot go into the details, but the three sub-modules show at the same time:

  - Differences with respect to the epistemic spheres, since two of them are based on the experimental data constituted at the level 2, whereas the third is a theoretical calculus.
  - And differences in content, since, for example, the two experimental sub-modules use two disjoint sets of machinic outputs.

  Let us assume that these differences are sufficient to see the three modules as *sufficiently independent*.[11]
- Fourth, the convergence of the three derivations appears to be of excellent quality: 9, 11 and 11 events, here are three results that nobody will hesitate to judge as very close to one another (two of them are even numerically identical). This uniformity of judgment is favored by the circumstance that the three outputs here involved are three *numbers*: results given in a quantified from, in general appear to be more easily and less problematically comparable than conclusions stated in a more qualitative form.

These characteristics of the 'Muon noise' module gives to its final result a strong degree of robustness, and justify the decision to give to this module (as analyzed just above) the status of a *prototypical example* of the general scheme (an *exemplar* in the Kuhnian sense). I think it is this type of example that commonly lurks behind the abstract scheme and feeds the intuitions about it.

10.13 A Revised Version of the Robustness Scheme

Now, a reflection on this example taken as an exemplar leads us to refine the first version of the robustness diagram that has been proposed at the beginning of this chapter (Fig. 10.2).

[11] As I developed in Chapter 1 (Section 1.8), the clause of independence is highly problematic. Clearly, in practice, independence is very often assumed intuitively and tacitly (without any systematic discussion or attempt of explicit clarification). The issue of independence is discussed further in this volume, notably in the contributions of Stegenga (Chapter 10), Nederbragt (Chapter 5) and Trizio (Chapter 4).

10.13.1 Recognizing the Gap Between the Multiple Sub-modular Conclusions and the Unique Totalizing Modular Conclusion

The first version suggested an identity of the conclusions at the output of each sub-module (represented by one and the same unique symbol '*R*'). Whereas the exemplar of the 'Muon noise' module clearly shows that, strictly speaking, we have three *distinct* results ($R_1 = 9$, $R_2 = 11$ and $R_3 = 11$).

This remark might seem merely anecdotal, especially when it is illustrated on such an example, in which two numerical values are identical and the third one is so close to the two others. But I think it is *not* anecdotal, and that in any case, it calls at least for an examination of the way we go from the individual outputs of the multiple sub-modules, to the unique output of the encompassing module.

Having obtained 9, 11 and 11, the value '10' is retained. This '10' is a unique totalizing estimation, built from the three sub-modular evaluations. How is it built? Actually, in the present case, we don't know. After giving the three results 9, 11 and 11, the experimenters immediately write without further explanation: "The correction to be applied is 0 ± 5 events". This being said, despite the absence of any explicit development of the matter, we can stress a number of contextual elements that act as constraints in the configuration under scrutiny.

First, the quantity to be estimated belongs to the category of a number of *events*, so it has to be an integer. Second, the estimated number is to be used to correct a definite number of NC- and CC-candidates. With respect to this aim, several different strategies are conceivable. For example: retain the highest number obtained by the different estimations, and examine if even the most pessimistic estimation (the maximal error) leads to a final corrected number of NC-candidates that is still sufficiently high to be interpreted in terms of the experimental detection of the NCs. Clearly, this is not the strategy that is retained here. Another possibility is to make the average of the three numerical values obtained and to round it off to the nearest integer: this would indeed lead to 10. Such a procedure would implicitly assume that each of the three derivations involved are equally reliable (and hence must be equally weighted). We can conjecture that it is what the experimenters did.

Anyway, the path the experimenters really followed in the present case is not important with respect to the general point I want to stress: namely, the *non-straightforward character of the equivalence* of on the one hand the three sub-modular values R_1, R_2 and R_3, and on the other hand the totalizing value R – a non equivalence which leads us to raise the question of the possibility that another path could have been followed.

In the passage from the three sub-modular values to the totalizing modular value, there is a jump. The passage involves a decision about the unique value R which will stand for the multiplicity of the three different values R_1, R_2 and R_3 obtained by the three different independent derivations L_1, L_2 and L_3. Very often – and this seems to be the case here –, the R is taken as the 'true value' (or the most-adequate-approximation-in-a-given-stage-of-knowledge, which in practice amounts to the same). In such a perspective, once the decision about the true value R has been taken, it feeds back on the epistemological status of the three intermediary conclusions: they become more or less close approximations of the true value: The ‘9’ must be corrected in ‘10’, the two ‘11’ also.[12]

10.13.2 History of Science, Individually Variable Reliability Judgments of Practitioners and the Contingency Issue

Such decisions depend both on (a) the history of science and (b) pragmatic intuitive evaluations of the individual scientists involved about the reliability of L_1, L_2 and L_3.

---

[12] Such claims involving a reference to a ‘true value’ are rarely made explicit, especially in published chapters, but I take them to be common intuitive ways of thinking among practitioners. In particular, such a framework commonly underlies the way practitioners understand and treat a series of actual measurement results associated with one and the same targeted variable when the results are obtained with one and the same instrument at different moments. Regarding this point, an interesting document is the 2008 version of the *International Vocabulary of Metrology – Basic and general concepts and associated terms* (VIM, 3rd edition, www.bipm.org). The document provides a unified vocabulary about “metrology, ‘the science of measurement and its application’ ” (p. vii), with the aim of being “a common reference for scientists and engineers (...) as well as for both teachers and practitioners involved in planning or performing measurements”, and “to promote global harmonization of terminology used in metrology” (p. 1), but it is of course not just a question of words. The elaboration of the final text has required an analysis of what it means to measure in the empirical sciences (an analysis of the different kinds of measurements, of the calibration procedures, of the basic principles governing quantities and units...), “taken for granted that there is no fundamental difference in the basic principles of measurement in physics, chemistry, laboratory medicine, biology, or engineering” (p. vii). Now in the final text, we read, in the introduction: “Development of this third edition of the VIM has raised some fundamental questions about different current philosophies and descriptions of measurement”. Two approaches are then contrasted: the “Error Approach (sometimes called Traditional Approach or True Value Approach)”; and the “Uncertainty Approach”. “The objective of measurement in the Error Approach is to determine an estimate of the true value that is as close as possible to that single true value. The deviation from the true value is composed of random and systematic errors. (...) [the two kinds of errors] combine to form the total error of any given measurement result, usually taken as the estimate.” (p. vii). It is not my aim here to explain the second approach, which is meant to get rid of the idea of a true value. I just want to stress that the first “traditional” approach coincides, in its fundamental features, with the one I have in mind in my analysis above. The second approach has been elaborated in response to the increased awareness that the traditional one – the only one involved in the previous versions of the VIM – was actually problematic (this point is still clearer in the 2004 first draft of the 3rd edition that has been submitted for comments and proposals to the eight organizations represented in the Joint Committee for Guides in Metrology (JCGM) and then revised according to their reactions; see the first paragraph of the Foreword). I take the fact that the “true value approach” has been the first one identified by the VIM, joined to the fact that it is subsequently described as the “traditional” one, as support in favor of the claim that it is indeed an intuitive, widespread largely tacit framework through which practitioners read the relation between different quantitative values obtained through different ‘derivations’ (in the VIM case: measurements) for one and the same targeted quantity. Since the 2008 edition is the result of a cooperation between numerous international experts, and since it has been approved by each of the eight member organizations of the JCGM, we can bet that it is representative enough with respect to claims about ‘widespread intuitive commitments’.

(a) Given the path of our history of science, it is today a quasi-automatic routine to use certain kinds of mathematical tools (especially statistical techniques) in order to interpret experimental machinic outputs (in order to evaluate the precision of instrumental devices; in order to build a unique measure from a series of measurements associated with different, more or less dispersed outcomes…). As a result of this historical path, the construction of one unique R through the operation of averaging a multiplicity of $R$is can hardly be seen today as a *creative* jump. Actually, it is even hard to be aware that there is any jump. However, there is one. To feel it, we have to go backward along the time axis and to realize how problematic and controversial it has been, historically, to legitimize and impose these mathematical techniques as the best ones.[13] The point can be generalized to any mathematical algorithm routinely and 'quasi-mindlessly' used in the empirical sciences today.

(b) Given the historical path and its crucial bifurcations, in a particular scientific context, the decisions relative to the construction of a unique R on the basis of a multiplicity of Ris depend on pragmatic evaluations of the practitioners involved in the research. In our case study for example, had the first derivation ($R_1 = 9$) been perceived as less reliable than the two other ones ($R_2 = R_3 = 11$), experimenters could have retained the value $R = 11$. Now, it is well known that judgments about what is reliable and what is not, or about the scale of what is more or less reliable, are a pragmatic (largely tacit) matter *often subject to individual variations.*[14]

[13] See for example (Bachelard 1927) and (Hacking 1990). As Bachelard stresses in his book, even the nowadays pervasive and obvious idea that an average value is a good way to represent a set of numbers has, historically, been the object of important discussions (see especially chapter VII). In a similar vein, see Buchwald (2006), a very interesting chapter on the developing methodology of taking statistical averages, from scientists in France and elsewhere, around 1800. I thank Thomas Nickles to turn my attention to that work.

[14] In the present historical episode, the existence of individual variations of judgments at many levels is largely documented and attested. See for example (Galison 1983, 1997; Rousset 1996; Schindler 201X). Moreover, with respect to pragmatic judgments about which derivations are reliable/unreliable, trustworthy/not trustworthy, and as a limiting case, worth mentioning or even considering as an argument, it has to be stressed that what appears in a published paper is the result of an antecedent invisible 'pre-selection' introduced by practitioners. In the paper, the reader finds three highly convergent derivations of the muon noise. But he knows nothing about other possible derivations elaborated during the investigation but finally put aside as 'unconvincing' and not mentioned in the paper. Moreover, all other things being equal concerning the reliability evaluations of the methods involved in a given derivation, the fact that the result of this derivation appears to be very far from the results of several others, can itself work – in cases where the 'non conformist' derivation under scrutiny is not based on already well-established approaches – as a reason to reject this derivation as unconvincing. This is another kind of operation through which the $R$is can be said to be 'mutually adjusted' in the course of the construction of a unique emergent R (in this case: because one of the $R$is does not fit with the others, it is eliminated – which means that the imputation of the discrepancy is directed toward its derivation: both the particular Ri and its derivation are discarded altogether as 'unreliable', 'too uncertain', etc.).

I do not claim at all to have shown by these brief considerations that the actual historical bifurcations or the actual options manifested in scientific published papers *could indeed have been different in an epistemologically significant way*. This is indeed a very hard philosophical issue, to which we can refer, following Ian Hacking, as the antagonism between "contingentism" and "inevitabilism" (see Hacking 1999, 2000). Even just a meaningful *formulation* of the issue would require too long a development to be provided here. Here I just want to suggest that contingentism should be taken seriously rather than being dismissed as too implausible from the very beginning without any true examination.[15]

10.13.3 An Act of Synthetic Calibrating Re-description

This being said, I hope the preceding reflections are sufficiently convincing to show that the final conclusion built as the output of the totalizing module (the value of R) must be considered, *in an important sense*, as a *different* and *new* conclusion with respect to each of the intermediary conclusions built as the outputs of the multiple modular components (the value of R_1, the value of R_2, etc.).

In the jump from the multiple $R$is to the unique R, one can say that there is *a certain kind* of mutual adjustment of the different intermediary results obtained as the outputs of the sub-modules. Had the experimenters opted for $R = 11$, the mutual adjustment would have been of another kind. In cases where R is viewed as 'the true value' and the $R$is as 'more or less close approximations of this true value', the decision to retain $R = 11$ rather than $R = 10$ leads to different feedback judgments with respect to the proximity of each of the $R$is to R and hence to different feedback judgments concerning the degree to which an Ri is a more or less good approximation. This can in turn have implications for the evaluation of the degree of precision of some instrumental devices or derivations. Suppose for example that the value '11' is taken as the true value instead of the 10: a subsequently introduced instrument or argumentative line that will lead to values centered on 10 will be considered, all other things being equal, as less precise or accurate than an instrument that will lead to values centered on 11.[16]

---

[15] Actually, I see this issue as the most fundamental and consequential issue of the philosophy of science and knowledge today. For a presentation and discussion of this issue and its philosophical implications, see Soler (2008a, b).

[16] This is just to sketch the general principle of some possible implications. In the present case, R is not meant to become an invariant physical quantity that will be subsequently measured by multiple instruments. Moreover, the kinds of implications just mentioned will only exist under the condition that the difference between $R = 11$ and $R = 9$ indeed makes a difference with respect to the aims of the investigation. In our example, this condition would primarily mean that the difference between 11 and 9 would engender, at the level of the two numbers of the NC-candidates left after the subtraction of the pseudos, a difference that would lead to cross the frontier between a 'yes' and a 'no' answer to the question of the experimental detection of the NCs.

What is the nature of the constitutive act involved in the passage from the *R*is to the *R*? By what kind of operation are the multiple sub-modular results converted in a single totalizing modular result? The move involves an operation that I will characterize as a (more or less *creative) calibrating (or standardizing) re-description*. Indeed, what do we do, from the sub-modular outputs *R*is to the totalizing modular output *R*? By an act of *synthesis* (which, depending on the situation, requires more or less ingenuity and creativity), we build, from the *R*is (here numbers, but they may be as well sentences expressed in more or less specialized words, mathematical equations, graphs, maps, pictures. . .), a new unique formula *R* (which can also take different forms) that is instituted as a *pole of reference* and *substituted for all of the Ris*. In the next stages, *R* will *stand for* the *R*is. The *R*is will be 'forgotten', and it is *R* that will be used as an unquestioned data in subsequent derivations (at the upper levels of the Gargamelle architecture). The act of synthesis involved operates a reduction of the manifold obtained at a certain level, and an identification of this manifold to one and the same thing, picked out by a new description, at another level. At the same time, it institutes the identity of this 'same' (assuming by doing so some operations of translation) and gives it the status of a pole of reference. In the frequent cases in which this pole of reference *R* is conceived as a 'true value' (as-we-know-it-in-the-present-stage-of-knowledge-of-course) rather than, for instance, a pessimistic threshold, *R* will work as a standard with respect to the precision of each *R*i (the more an *R*i will be far from *R*, the less this *R*i and the argumentative line from which it has been derived will appear precise).

Admitting the preceding reflections, we are led to introduce some modifications to the first version (Fig. 10.2) of the robustness diagram. No longer do we have three arrows all ending at one and the same point, the result *R*, but (see Fig. 10.11) three arrows each ending at three different results, themselves then synthesized in a unique result at an upper emergent level. Or alternatively, relying on the modular representation (see Fig. 10.12) three sub-modules, with three outputs R_1, R_2 and R_3; one inclusive module with an output *R*; and an intermediary space between the horizontal of the Ris and the horizontal of the Rs, the depth of which represents the importance of the creative jump.[17] I will characterize the structural fragment

[17] As we saw, the jump can appear more or less creative according to the case. In the case of the 'Muon noise' module, the jump would certainly not be perceived as creative by anybody. Actually, it is even difficult to see that there is any jump. This could lead to discard this example as a good means to give credit to the general epistemological point at stake, namely that the *R* must be considered as a significantly different result with respect to the *R*is. Indeed, I concede that other examples would help to understand better and reinforce the point (see just below for references to examples in the present book).

But my strategy has precisely been to show that even in cases in which the passage from the *R*is to the emergent *R* might seem to be completely automatic and uniquely imposed, this passage nevertheless involves operations of conversion that have to be recognized. Indeed, the fact that they are not recognized – or in other words the reference to configurations in which they are almost invisible like in the example of the 'Muon noise' module – is precisely what fuels the no-miracle argument that lurks behind the robustness scheme of the three arrows converging on one unique *R* and makes it appear so convincing (see Section 10.3). Maybe, in some historical cases, we can talk

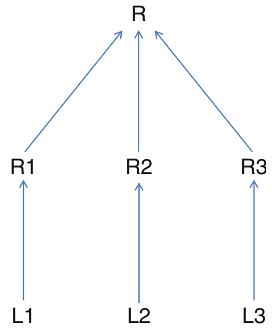

Fig. 10.11 A revised version of the robustness scheme: The elementary scheme of robustness in the *arrow-node* representation

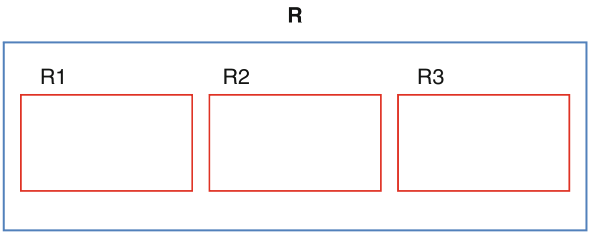

Fig. 10.12 A revised version of the robustness scheme: The elementary scheme of robustness in the *modular* representation

represented by this figure as the *elementary scheme of robustness*. The Muon noise module is a prototypical exemplification the elementary scheme.

In order to understand in what sense this scheme can be said 'elementary', let us come back to the global architecture of the Gargamelle experimental argumentative line.

as if the operations of conversion involved are not significant (are indifferent with respect to certain aims). But in order to be in a position to draw this conclusion legitimately, we have first of all to recognize the very existence of such operations and examine the kind of work they accomplish in each case.

For other examples developed in this volume which could help to understand better and reinforce the point here put forward, see notably the Chapter 9 of Stegenga Section 9.4 (about the transmission of a virus in epidemiology), the contribution of Trizio Chapter 4, Section 4.3 and the article of Allamel-Raffin and Gangloff Chapter 7, Section 7.6 (about the production of maps in astronomy). Through the analyses proposed in the latter article in particular, we clearly see how the images first obtained with different kinds of telescopes have to be manipulated and transformed before they can appear 'essentially similar' one to the others. As a result, a new and still questionable couple 'derivation-result' is viewed to be 'in essential agreement' with more ancient and already taken-as-established ones. This harmony works as an argument in favor of the new derivation-result couple under discussion. Because the new result is seen as 'the same' as already taken-as-robust old ones, it follows that, jointly, the new derivation is taken as solid and the new derived result is taken as robust. Once this has been achieved, the situation is re-described as: multiple derivations lead to one and the same invariant result. But as soon as we examine the details of the historical process, we find that the new result and the ancient ones were not immediately 'the same' from the start. The 'initial' images indeed have been transformed, through certain specifiable operations, in order to become comparable.

10.14 The Elementary Scheme of Robustness and the Global Architecture of a Derivation

10.14.1 Identifying the Elementary Schemes of Robustness Inside the Gargamelle Modular Architecture

Inside an architecture of the Gargamelle type we find, locally, some modular units that satisfy a scheme *akin to* what I just called the *elementary prototypical* scheme of robustness. I don't have the space to justify this claim. To justify it, we would have to:

- First, describe the content of the modules that can pretend to be akin to the prototypical elementary fragment.
- Second, analyze the differences with respect to the prototype exemplified by the 'Muon noise' module. For example:
  - Quantitative versus qualitative conclusions;
  - The more or less creative and more or less problematic character of the act involved in the passage from the sub-modular to the modular conclusions;
  - The involvement of parallel procedures in which the *similarities* largely dominates the differences[18]...
- Third, discuss the way in which these differences with respect to the prototype can influence the feeling of 'miracle' (see above Section 10.3) and thus the robustness associated with the totalizing result.
- Fourth and finally, argue the decision that *despite the differences* involved, we are entitled to assimilate the modules in question to *variants* of the prototypical scheme of robustness.

[18] Regarding this point, the example of the 1974 paper about NC is interesting. In my account above, I simplified the presentation in many respects. One of these is that the actual analyses of the 1974 article are in fact constituted of *two* parallel investigations: the one devoted to the neutrino-induced interaction (on which I have exclusively focused above), and *another one* devoted to the similar *anti*-neutrino-induced interaction (i.e., an anti-neutrino interacts with a nucleon, leading to a positive muon and a shower of hadrons). Essentially the same treatment is applied to one and the other case (or in other words: the anti-neutrino case can be reconstructed through essentially the same structural architecture than the one I sketched for the neutrino case). No doubt, the harmony obtained, in terms of the totalizing outputs of the different modules on each floor of the architecture, between the neutrino case on the one hand and the anti-neutrino case on the other hand, also contributes, to the practitioners' eyes, to reinforce the confidence in the final conclusion of the chapter in favor of the plausibility of weak neutral currents. Now, this reinforcement configuration, primarily based on treatment *similarities* (by opposition to *different, independent* treatments), does not correspond to a Wimsattian robustness scheme.

I will content myself to make visually apparent, within the global architecture, *some* of the modules that are, in my opinion, good candidates to the title of variants of the elementary prototypical scheme of robustness (see Figs. 10.13, 10.14, and 10.15).

10.14.2 *Fractal Articulations of Robust Elementary Units and Other Kinds of Articulations*

How are these minimal fragments of robustness, these locally robust blocks, involved in the overall construction?

First it is remarkable that, inside a floor, one often finds a sequence of elementary schemes of robustness included one inside the other, that deploy themselves in a kind of spiral on several successive contiguous level of emergence. Indeed, let us start from the deeper elementary fragment of robustness, say the fragment of level *N* (see, on Fig. 10.13, the colored module of the second floor). Once constituted, the totalizing output of this module acquires autonomy with respect to the derivations involved in the multiple parallel sub-modules of level *N* that have made it robust, and it is then used as the input of one of the parallel sub-modules that constitutes a new elementary pattern of robustness at an immediately superior level of emergence *N*+1 (see, on Fig. 10.14, the colored area at the second floor)... And so on from level to level (see Fig. 10.15 second floor). Something like a fractal is involved, the elementary scheme of robustness being the minimal pattern that is repeated at each level.

But the architecture is not entirely constituted in this way, following such a fractal algorithm.

First, the building blocks are not all structures that resemble the elementary scheme of robustness. Consider for example each of the sub-modules 'Selections'. If it is often possible to give several parallel arguments in favor of the output of one of these modules, this output does not seem to be retained because it appears as the stable point of convergence to which the different parallel arguments end.

Second, even when the multiple sub-modules of a parallel set satisfy the pattern of the elementary scheme of robustness, their totalization in terms of an encompassing module does not always follow the same scheme. The operation by which one goes from the sub-modular outputs to the totalizing modular output is not always a synthesis of the calibrating re-description type. Sometimes, for example, the sub-modular parallel outputs simply add up. This is the case for the module 'Noises' (second floor). Or sometimes, the sub-modular outputs are in a relation of mutual exclusion. This is the case for the module 'Theoretical interpretations' (third floor).

On the whole, in the internal architecture of an argumentative line of the Gargamelle type, one finds local fragments of robustness and fractal articulations of such fragments, *but not only that*. The modules that satisfy the robustness scheme are embedded in a complex network and articulated through different combinations with other types of modules.

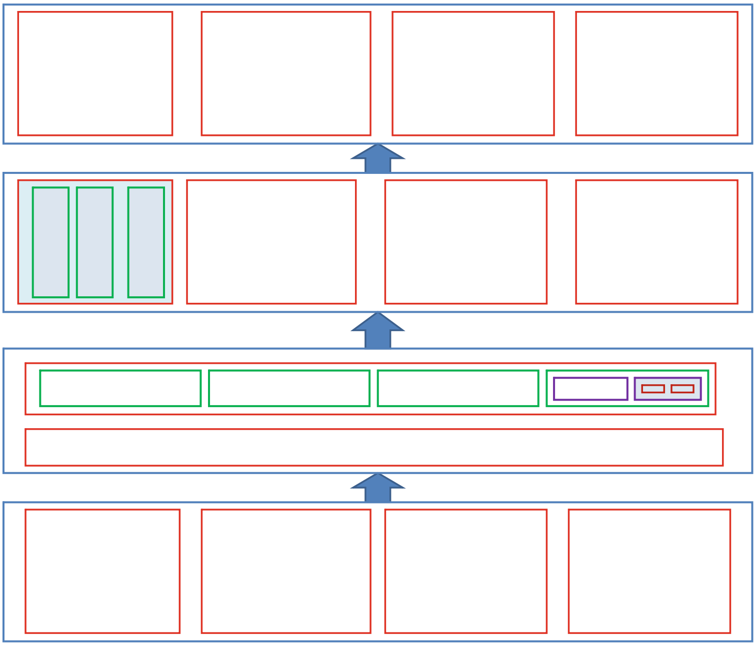

Fig. 10.13 Localisation of the elementary schemes of robustness inside the Gargamelle architecture (1)

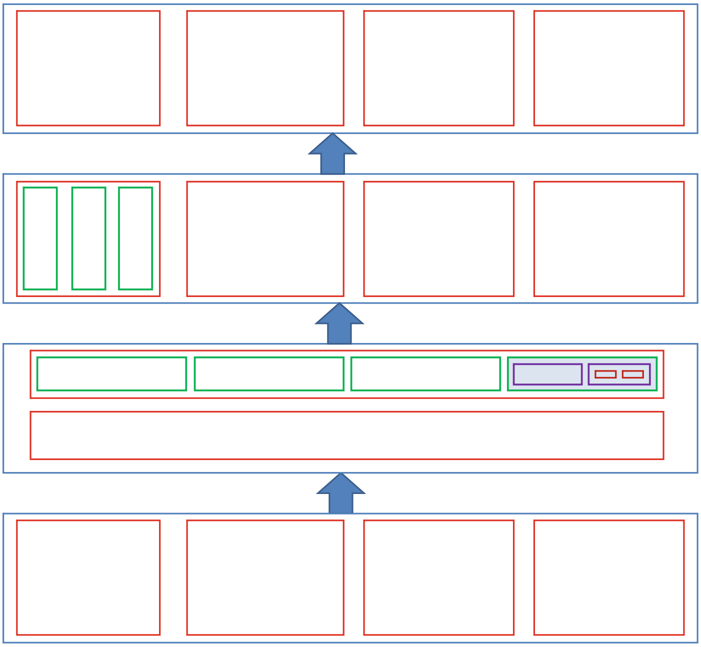

Fig. 10.14 Localisation of the elementary schemes of robustness inside of the Gargamelle architecture (2)

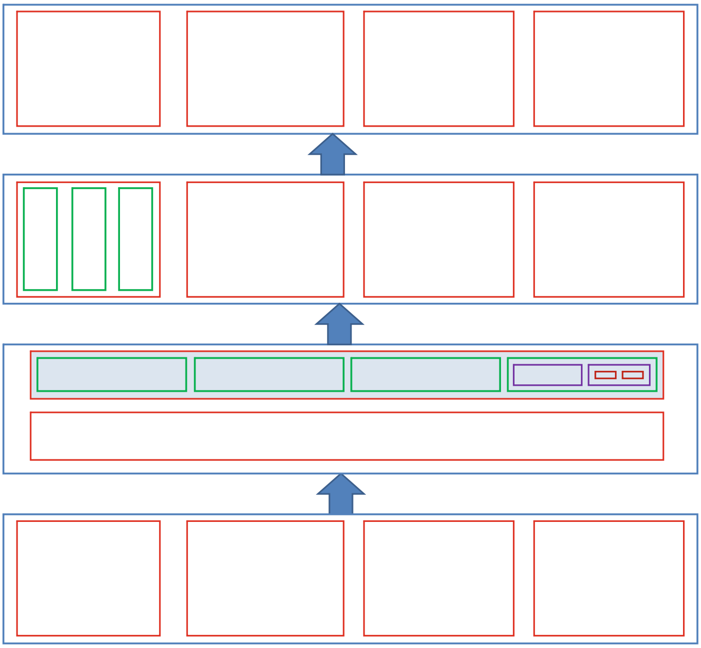

Fig. 10.15 Localisation of the elementary schemes of robustness inside of the Gargamelle architecture (3)

10.15 The Solidity of a Derivation Considered as One Derivational Unit: The 'Internal' and 'External' Contributions to Solidity Attributions

All this being admitted, to what is finally due the solidity of a derivation *considered as a whole* (for example *the* Gargamelle experimental line)?

- One side of the answer relates the solidity of the derivation as a whole to internal characters: to what the derivation is intrinsically made of.

  On that side, solidity analysis will be based on a characterization similar to the one I just proposed about the Gargamelle line. And admitting the characterization I have proposed, the solidity of the line as a whole must be referred, not *simply* to the elementary scheme of convergence under multi-determinations, but to a *much more complex* scheme inside of which the elementary scheme is involved *as an ingredient*. In such conditions, the solidity of the whole (of *the* Gargamelle arrow-derivation) owes a lot, not only to the degree of robustness of those of its parts that satisfy the robustness scheme, but also to the repertory of all the sub-modules involved, to the structure and content of each, as well as to the manner they are combined with one another. The solidity of the whole is related to a 'global fit' between the multiple ingredients involved in the architecture (see Soler 201X).

- The second side of the answer relates the solidity of the argumentative line as a whole to external or extrinsic circumstances: to the existence of *other* different convergent derivations.

  On that side, one will link the robustness of the Gargamelle line to the fact that a second type of experimental argument exists – for example an electronic one – and one is led to the same conclusion as the visual Gargamelle argument. Here, the solidity is asserted on the basis of the elementary scheme of robustness itself (and not on the basis of a more complex scheme).

Taking the two sides of the answer together, the solidity of an experimental argument of the Gargamelle type can be investigated:

- Either by looking inside the procedural black box (inside the Gargamelle argumentative module) – and in this case the solidity will be related to its internal constitution;
- Or by looking outside the procedural black box (outside the Gargamelle argumentative module), i.e. by examining the role played in more extended networks by the argumentative line treated as one unitary black box – and in this case, the solidity will be related to the situation of this line with respect to different argumentative lines (and often to the fact that the derivation under scrutiny is involved as an arrow in an elementary scheme of robustness).

From an analytic point of view, it seems desirable to distinguish these two parts of the answer. This being said, in a given historical situation, the solidity of a derivational modular *unit* considered *at a given scale* (for example the Gargamelle derivation as we defined it) can be due, *either* mainly to what is *inside of it*, *or* to what is *outside* of it, or to *both* what is inside and outside. The respective contributions of the inside and outside configurations to the solidity of the argumentative line, and the directions of the 'solidity fluxes' (from the inside to the outside or from the outside to the inside), will depend on the historical path. Or more exactly, they will depend on the 'solidity values' that are initially attributed by practitioners to the multiple ingredients involved in the historical situation (which will in turn depend on the past history of science: on what is, given this history, taken as already firmly established/discussable, reliable/not so reliable and so on. See Chapter 1 (Section 1.7)). For example, in case an electronic experimental argument is led to the same result as a previously obtained visual argument first viewed as fragile taken in isolation, this will reinforce the overall visual experimental argument considered as a whole as well as its multiple ingredients. Here, the solidity flux will be mostly directed from the outside to the inside. But in case the internal ingredients of the visual argument are already taken from the beginning as especially solid, this can be enough to judge that the visual argumentative module taken as a whole is sufficiently solid 'in itself' (independently of any other 'external' convergent derivation).[19]

[19] It is worth noting that such a characterization of the situation is a conceptualization of the analyst (the historian or the philosopher of science), and a reconstruction which, with respect to its

10.16 Conclusions

10.16.1 What the Robustness Scheme Provides and Cannot Provide with Respect to the Analysis of Scientific Practices

The scheme of robustness 'convergent results under multiple independent derivations' is useful, and even indispensable, in order to describe science. I think that usually, when a philosopher of science asks the question of how a taken-to-be-robust scientific achievement has acquired this status historically according to scientists, he will be able to give an explanation involving an elementary scheme of robustness.

But at the same time, it is important to stress that the scheme is just *a part* of the explanation, in the sense that the structure of its skeleton is insufficient to account for the positions and decisions of living scientists (and *a fortiori* of no help for anticipations of practitioners' future options from a given scientific stage characterized by a scientific debate). Indeed, this structure in itself tells us nothing, in a given situation:

- Neither about the number of independent convergent derivations required to conclude to a sufficient robustness;
- Nor about the required degree of independence of the parallel lines;
- Nor about how to estimate the force of each of the parallel derivations that play in favor of a result;
- Nor, finally, on how to weight the different parallel argumentative lines in the cases (historically frequent) in which only some of them converge *but others disagree*[20]...

faithfulness to real cases, might be difficult to establish and remains highly conjectural by nature. This is because crucial ingredients of the story which are supposed to determine the 'fluxes of solidity', such as "the 'solidity values' that are initially attributed by practitioners", "an argument first viewed as fragile taken in isolation", "an argument already taken from the beginning as especially solid" and other appreciations of this kind, most of the time remain *tacit and opaque* to practitioners themselves (this is what I called the opacity of experimental practices with respect to description and justification). See Soler (2011). Hence they are not the kind of things to which the analyst has a transparent and unproblematic access. Actually, even their very existence can be questioned. It is discussable that these kinds of ingredients, postulated by the analyst in order to make a historical episode more understandable, can be equated to well-determined empirical facts (typically to well-defined stable states of the mind of each real subject of knowledge, or to a collective 'tacit basis' shared by the members of a scientific community). In any case, practitioners are usually not well-aware of such states, and when the sociologist of science asks them questions about their confidence in this or that ingredient of their science, the answers are not something like unproblematic numbers or unambiguous sentences that the analyst could immediately identify as such, without any discussion, to the 'solidity values' according to which real practitioners indeed rated, in the actual historical sequence, the different ingredients involved.

[20] This is congruent with Stegenga's conclusion in this volume: "robustness-style arguments do not tell us what to believe in situations of evidential discordance."

Of course, the (more or less explicit) positions that practitioners will endorse with respect to these intertwined points will depend, not on the scheme of robustness as a form or a skeleton, but on the scheme *as fed with a certain specific content*. It seems, thus, that the scheme *as a structure* is not what counts primarily. In any case, the scheme as a structure is not a sufficient condition to impose in a compelling, uniform way, the robustness (or a determined 'degree of robustness') of the node *R* involved in it. It is a form that the philosopher indeed finds when he analyses science, but once such a form has been exhibited in a particular case, it is not enough to 'justify' or 'explain' the practitioner's judgments that *R* is robust (or sufficiently robust). To account for such judgments (as far as this can be done), the philosopher will have to take into account the particular content to which the robustness skeleton is associated in each case.

This is indeed not a surprising conclusion, if we take into account the teachings of Science Studies devoted to the chapter on 'scientific method' in recent decades. As Thomas Kuhn showed already in the 1960s, and as the concept of scientific paradigm was intended to stress, practitioners' judgments about what is reliable/unreliable, trustworthy or not, scientific or metaphysic, etc., have an irreducible *pragmatic* dimension (see Kuhn (1970), and Kuhn (1973) on values in scientific judgments). This means, among other things, that they are not reducible to the inescapable output of an algorithmic calculus that the philosopher of science could make entirely explicit, and in which the different ingredients of the historical scientific configuration under scrutiny would be uniformly and univocally weighted. Actually, the substance of such judgments remains opaque, and the explicit positions of different practitioners about what is reliable and what is not often appear to be divergent. There is no reason why robustness attributions should be an exception.

Taking all that into account, we should not ask too much of the robustness scheme. Once this is recognized, the robustness scheme is indeed a very useful analytic tool in order to analyze actual scientific practices. Indeed, the general scheme exhibits and characterizes a central and pervasive pattern that underlies practitioners' judgments about the quality of scientific achievements, and this helps to recognize instantiations of the general structure in particular historical cases and to clarify its specific substance in each case.

In the perspective of this kind of 'weak program' about scientific method, the contribution of the present chapter has been to provide a reflection on what makes the solidity of the arrow-derivations involved in a Wimsattian robustness scheme. A given derivation can borrow its solidity, both from 'external factors' (namely from its position as an arrow in a robustness configuration in which the other arrows and nodes are already taken-as-sufficiently-solid) and from its 'internal' features. On this second side, the chapter has shown that the solidity of the derivation as a whole can be analyzed as a *global good fit* of a *much more complex structure* than the one of the elementary scheme of robustness (a complex structure involving multiple elementary schemes of robustness as ingredients). More work would be required to characterize the nature of such kind of complex fit and the kind of glue(s) involved in it. But on the basis of the present reflection, we can suggest that the robustness scheme is *one particular*, indeed especially prominent, kind of holistic fit *among* *other possible ones*, through which something might acquire the status of a solid achievement in the course of the history of science.

10.16.2 Epistemological Open Issues and Lines of Future Investigations

The analyses of this chapter are not devoid of consequences as regards issues of epistemological significance, such as the traditional one of scientific realism, and the less traditional one introduced above (Section 10.13) as the antagonism between contingentism and inevitabilism. In my opinion, the above analyses raise doubts about the plausibility of scientific realism and inevitabilism.[21] To close this chapter, I would like to indicate why I think the preceding analyses weaken correspondence realism and realist-inspired inevitabilism, by sketching the kinds of arguments they suggest (for more details, see Soler (201X)).

When we scrutinize what is behind robustness judgments, and in particular when we analyze what each of the multiple derivations is made of, the upshot encourages us to be extremely cautious about the passage from the robustness scheme to inevitabilism and scientific realism.

First, we should keep in mind two possible ways of speaking about a robustness scheme, which at first sight can both appear equally legitimate, if not quasi-equivalent, but which actually reflect and generate *strongly different intuitions*, and thus surreptitiously act as supportive elements for or against epistemological stances akin to realism and inevitabilism.

In order to describe the passage of the multiple sub-modular outputs to the unique totalizing modular output, we can say – and we commonly say – that the parallel derivations lead to an invariant something, or converge on one and the same result. This is certainly not false. But this formulation suggests an invariance *already given as such* at the level of each sub-modular outputs, a *pre-determined ineluctable identity* that scientists have bumped into, much *as we bump against a wall*. This image strongly pushes us toward reading of the convergence in terms of the 'no-miracle argument' (see Section 10.3 above), and hence fuels the realist-inspired inevitabilist conviction.

Whereas if, in order to stress the creative act of synthesis involved in what I called the calibrating re-description, we depict the passage from the multiple sub-modular outputs to the unique higher-level modular output through a formulation of the kind 'the different parallel derivations have led to strictly speaking different conclusions that practitioners did succeed to conciliate by substituting to all of them a unique

[21] Realist and inevitabilist commitments, although analytically distinguishable, very often go hand in hand concretely, since most of the time, inevitabilism is endorsed as a result of a realist stance: Such or such ingredient of our science is thought to be inevitable because it is taken as a faithful description of a bit of a unique world which is what it is once for all independently of scientists. See Soler (2008a, Section 3).

totalizing calibrating conclusion', the flavor is rather different... In particular, the feeling that the convergence would have to be seen as a 'miracle' – i.e. it would remain *completely unexplained – without an invocation of the pressure of 'reality'*, this feeling is strongly attenuated. No doubt, it is certainly very hard, and sometimes not convincingly possible, to make a multiplicity of new results cohere with one another so that all of them moreover nicely fit with the extended stock of the other already taken-for-granted scientific achievements. But if there is a 'miracle' here, it seems to be of a different kind than the one involved in the realist argument (on this point, see also Chapter 1, Section 1.8).

The relevance of each of these two possible re-descriptions of the robustness scheme has to be estimated case by case. But the very possibility that the first, usual formulation could actually hide a situation of the second kind, at least encourages us to be extremely cautious with respect to the quasi-irresistible tendency, undeniably active for practitioners and for each of us in ordinary situations, to assimilate what is robust to what is true, to what reveals a bit of reality and hence was inevitable (given, of course, some – admittedly partially contingent – 'initial' historical conditions, such as the questions scientists asked, the instrumental means at their disposal and so on).

Second, another complementary line of reflections also raises doubts on the intuitive realist reading of the robustness scheme.

The way the abstract scheme 'multiple arrows converging on one result *R*' instantiates in a given historical situation is strongly dependent on a conceptual and theoretical shaping: ways of analyzing problems, of elaborating questions, of deciding about the relevant variables and strategies (and as a particular case: of building this calibrating re-description)... Each module, each modular decomposition and global architecture, comes into existence on the basis of such a shaping. Now it is difficult to argue that such a shaping is, as such, written in nature or even uniquely imposed by what is already taken as the 'scientific facts' in a given stage of knowledge. Yet to be in a position to argue that something like that holds, *at one stage* of the investigation *or another*, is needed in order to support correspondence realism and inevitabilism. Otherwise, if several ontologically disparate solid fits are at all stages convincingly possible, if there is *no point at which one is uniquely imposed*, we are led to the idea of an alternative science that could be, at the same time, *both solid* (in the same intuitive sense we say our science is solid) *but ontologically very different* from our science. In other words, we are led to a contingentist position.

Let me say a little bit more about this point, starting from the shaping that lurks behind the modular architectures on the basis of which an item acquires its status of established result.

First of all, it has to be stressed that according to the context, the constitutive act of shaping involved in a modular decomposition appears *more or less* creative and problematic.

Sometimes *its very existence* can remain invisible to practitioners. This is the case when a given modular decomposition appears, in a given stage of scientific practices, almost automatic, obvious, deprived of any alternative and hence strongly compelling. In our historical episode, this is the case for the module 'Collective treatment' of the second floor: the division of the initial problem into three parallel investigations which respectively focused on the spatial, on the energetic and on the angular distributions was, at the time, a usual and almost inescapable step of the interpretative practices of the photographs obtained with visual detectors in high energy physics.

In some other contexts, the modular decomposition appears optional, creative and potentially problematic to the actors themselves. For instance, in the 1974 paper, the decision to investigate the NC/CC *ratio* (rather than to study the NC *considered in isolation*) appears as an *optional strategy* (although it is completely constitutive of the final conclusions since, as we have seen, the CC works as an experimental standard for the identification of the NC and for their differentiation from the neutron background[22]). Correlatively, the repertory and the treatment of the different kinds of relevant background events is quite problematic (the multiple practitioners involved in the research about weak neutral currents at the time were not worried about the same risk of confusion; they did not trust the same kinds of methods for the evaluation of the noises; they were deeply aware that some still unidentified pseudo could have been missed...). The treatment of the most problematic background event, the so-called "neutron background", is itself a highly complex architecture inside the Gargamelle construction, and its modular constitution involves some rather creative steps (through the use of a Monte Carlo simulation, itself based on multiplicity of uncertain hypotheses about the properties of neutrons and neutron cascades).

Now, whether perceived as optional or inescapable, creative or imposed, problematic or obvious, it is very difficult to see something like a modular structure as inescapable or 'uniquely imposed'. Or more exactly, when practitioners have the felling of inescapability, it is on the basis of an anterior historical trajectory which is itself made of similar modular structures. And so on indefinitely: we never find anything else than modular structures, past or present (see Chapter 1, Sections 1.6.2 and 1.7.2). Now a modular structure – and as a particular case a robustness scheme – works as a holistic equilibrium, and a great deal of contingencies are involved in the constitution and emergent conclusions of a holistic equilibrium.

The acts that determine the number and content of the parallel modular units, as well as the modalities of their articulations in more complex architectures involving different levels of emergence, are, uncontroversially, dependent on a partially contingent history. For example, if the CCs can be instituted as experimental standards with respect to the identification of the NCs, this is as the result of the – in themselves uncontroversially contingent – programmatic choices made in the previous years (i.e. the choice to conduct experimental studies of the CCs instead of other sub-atomic phenomena). At first sight this obvious remark seems anecdotal, but its potential harmfulness appears when we stress that *what has been done* and *what has been established* (until-further-notice-of-course) *in the past*, is *not at all indifferent*, and *strongly conditions*, what is done and what is taken as established *in the future*. So a genuine path-dependency could be at stake.

[22] On this constitutive role, see Pickering (1984) for more elements.

In each 'synchronic' stage of the history of science, what is taken to be plausible or established acquires its status from its situation inside of a global equilibrium structurally similar (although much more complex) to the Gargamelle architecture of our example. At each floor of this architecture, a module works as a holistic equilibrium. Imagine a change at one point or another: it is plausible that the emergent conclusions would have been different. With different or additional derivations at the level of a given module, the global assessment (the totalizing output) could have been different.[23] Now what is available and what is not in terms of derivations depends on the past history. And what is built as the output of a given floor is not indifferent to what is built at the immediately superior floor. Indeed, what is obtained at a lower level as the result of a local equilibrium (*this or that* totalizing output), is, subsequently, used as an unquestioned, given *datum* (plays the role of input) for the constitution of new equilibriums at higher levels of emergence. So that we can talk about a sort of *amplification or cascade process* when we jump from a level of the Gargamelle architecture to the next one. And arguably, something structurally similar holds when we consider the diachronic relations between successive synchronic 'slices' of scientific developments.

I do not claim to have demonstrated contingentism against realism and inevitabilism. As already indicated, my aim has only been to sketch the directions in which a genuine argument would have to be sought. The core of the argument would lie in the way human knowledge is built, that is, as a succession of holistic "symbioses" (in Pickering's terminology) resting one on the other along a diachronic line. Reflecting on such structural characters, it becomes difficult to assert that at a point or another (be it at the 'ideal end of research'), the contingencies related to the way the problems have been framed at each level, the contingencies related to way the solutions have been constructed as the converging points of multiple available derivations (or through more complex global good fits), or in brief, the contingencies of the whole process of the deployment of successive modular arborescences and equilibriums in the course of the history, can be erased, eliminated so as to impose a *unique* story, let alone a *true* unique story that could pretend to mirror (or at least to map in an isomorphic manner) a unique physical world which is what it is once and for all. It is in that sense that a reflection on the working of a modular architecture of the Gargamelle type weakens the intuitive obviousness of the inference from a robustness scheme to correspondence realism and inevitabilism.[24]

---

[23] This is perhaps even possible on the basis of fixed derivations (although this will certainly appear more questionable), if we admit that the totalizing unique output is a calibrating re-description that is not uniquely imposed by the multiple results obtained as the outputs of the lower-levels sub-modules.

[24] Thomas Nickles' chapter – through an analysis which, although inspired by a quite different literature from the one which inspired my own chapter, also seeks to study the implications of the structural features of the humanly designed epistemic systems – draws congruent conclusions about correspondence realism and contingentism.

Acknowledgements I am especially indebted to Emiliano Trizio for multiple helpful feedbacks and extended discussions about the claims of this chapter. I am also grateful to Cathy Dufour, Thomas Nickles and Jacob Stegenga for their useful comments. Finally, many thanks to T. Nickles, J. Stegenga and E. Trizio for their corrections and suggestions of improvement concerning the English language. The end result is, of course, my own responsibility!

References

Allamel-Raffin, Catherine. 2005. "De l'intersubjectivité à l'interinstrumentalité. L'exemple de la physique des surfaces." *Philosophia Scientiae* 9(1):3–31.

Bachelard, Gaston. 1927. *Essai sur la connaissance approchée*. Vrin; 6th ed., 1987.

Benvenutti, A. 1974. "Observation of Muonless Neutrino-Induced Inelastic Interactions." *Physical Review Letters* 32(14):800–3.

Buchwald, Jed Z. 2006. "Discrepant Measurements and Experimental Knowledge in the Early Modern Era." *Archives for History of Exact Sciences* 61:1–85.

Galison, Peter. 1983. "How the First Neutral-Current Experiments Ended." *Review of Modern Physics* 55(2):477–509.

Galison, Peter. 1997. *Image and Logic, A Material Culture of Microphysics*. Chicago: The University of Chicago Press.

Hacking, Ian. 1990. *The Taming of Chance (Ideas in Context)*. Paperback.

Hacking, Ian. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, edited by A. Pickering, 29–64. Chicago and London: The University of Chicago Press.

Hacking, Ian. 1999. *The Social Construction of What*? Cambridge, MA: Harvard University Press.

Hacking, Ian. 2000. "How Inevitable are the Results of Successful Science?" *Philosophy of Science* 67:58–71.

Hasert, F.J. et al. 1973a. "Search for Elastic Muon-Neutrino Electron Scattering." *Physical Letters* 46B:121–4.

Hasert, F.J. et al. 1973b. "Observation of Neutrino-Like Interactions Without Muon or Electron in the Gargamelle Neutrino Experiment." *Physical Letters* 46B:138–40.

Hasert, F.J. et al. 1974. "Observation of Neutrino-Like Interactions Without Muon or Electron in the Gargamelle Neutrino Experiment." *Nuclear Physics* B 73:1–22.

Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: The University of Chicago Press.

Kuhn, Thomas. 1973. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension*, Selected Studies in Scientific Tradition and Change, 320–39. The University of Chicago Press, 1977.

Pickering, A. 1984. *Constructing Quarks, a Sociological History of Particle Physics*. Chicago and London: The University of Chicago Press.

Rousset, André. 1996. *Gargamelle et les Courants Neutres, Témoignage Sur Une Découverte Scientifique*. Presses de l'Ecole des Mines de Paris.

Soler, Léna. 2008a. "Are the Results of Our Science Contingent or Inevitable? Introduction of a Symposium Devoted to the Contingency Issue." *Studies in History and Philosophy of Science* 39:221–29.

Soler, Léna. 2008b. "Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue." *Studies in History and Philosophy of Science* 39:230–41.

Soler, Léna. 2008c. "The Incommensurability of Experimental Practices: The Incommensurability *of What*? An Incommensurability *of the third-type*?" In *Rethinking Scientific Change and Theory Comparison. Stabilities, Ruptures, Incommensurabilities?* edited by L. Soler, H. Sankey, and P. Hoyningen, 299–340. Springer, Boston Studies for Philosophy of Science.

Soler, Léna. 2011. "Tacit Aspects of Experimental Practices: Analytical Tools and Epistemological Consequences." *European Journal for the Philosophy of Science (EJPS)* 1(3):394–433 (Spécial issue directed by the Society for Philosophy of Science in Practice, Mieke Boon, Hasok Chang, Rachel Ankeny et Marcel Boumans).

Soler, Léna. 201X. "A General Structural Argument in Favor of the Contingency of Scientific Results." In *Science as it Could Have Been. Discussing the Contingent/Inevitable Aspects of Scientific Practices*, edited by Léna Soler, Emiliano Trizio, and Andrew Pickering. In progress.

Wimsatt, William. 1981. "Robustness, Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins, 125–63. San Francisco, CA: Jossey-Bass. Reprinted in *Re-Engineering Philosophy for Limited Beings, Piecewise Approximations to Reality*, 43–71. Cambridge, MA, and London, England: Harvard University Press, 2007. Reprinted in this volume, Chapter 2.

Chapter 11
Multiple Means of Determination and Multiple Constraints of Construction: Robustness and Strategies for Modeling Macromolecular Objects

Frédéric Wieber

11.1 Introduction

As Wimsatt has repeatedly and explicitly pointed out,[1] one of the main sources of his interest in the concept of robustness and in the procedures of robustness analysis is constituted by the work of Levins (1966, 1968) on "robust theorems" in the context of model building in population biology. If Wimsatt has expanded the notion of robustness to a much broader set of procedures and contexts of use than what was intended for by Levins' discussion,[2] the work of Levins still seems to be a good resource to discuss robustness in model-based scientific activities. More generally, the overall discussion within which the notion of robustness analysis is introduced in Levins (1966) is of great interest when discussing practices of modeling (Weisberg 2006a).

It is within a discussion of the practical limitations population biologists face when constructing and analyzing models that Levins (1966) has emphasized the interest of robustness analysis. Because the systems encountered in population biology are complex, the construction of *manageable* models of these systems requires *tradeoffs*. As Odenbaugh (2006) has convincingly argued, "(. . .) Levins' discussion of tradeoffs in biological modeling concerns the tension between our own limitations with respect to what we can compute, measure, and understand, the aims we bring to our science, and the complexity of the systems themselves" (Odenbaugh 2006, p. 618). After having emphasized such specific constraints of the

[1] See Wimsatt (1981), for example p. 124 or p. 126. See also Wimsatt (1994) or (2001).

[2] Wimsatt (1981) wrote: "The family of criteria and procedures which I seek to describe in their various uses might be called *robustness analysis*. (. . .) I will call things which are invariant under this analysis 'robust', extending the usage of Levins (1966, p. 423), who first introduced me to the term and idea and who, after Campbell, has probably contributed most to its analysis" (p. 126).

F. Wieber (✉)
Archives H. Poincaré, Laboratoire d'Histoire des Sciences et de Philosophie, UMR 7117 CNRS, Nancy, France
e-mail: frederic.wieber@wanadoo.fr

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_11,

epistemological context in population biology, Levins (1966) analyzes the modeling practices of population biologists and considers that three different modeling strategies, implying different tradeoffs, are developed in this field. One of these strategies is the one Levins uses and prefers, as a modeler. It is within this strategy, where "very flexible models" (as Levins 1966, p. 19, calls them) are constructed, that robustness analysis emerges.

In this chapter, I will present and analyze a theoretical procedure developed during the 1960s and 1970s within the field of protein chemistry. The interest of this case study is definitely linked with such a picture of science interested by the limitations scientists encounter and the "in practice" strategies they then devise in order to study "messy systems" (a wimsattian expression[3]). So, it will be necessary to grasp the epistemological context within which such theoretical procedure, which constitutes a type of protein modeling, has been set up by scientists: what are the specific problems scientists were faced with when constructing this procedure? For theoretical chemists, proteins are truly complex objects, and the construction of a theoretical method to gain knowledge on the properties of these objects by devising models of their structures faces them with different kinds of limitations (notably computational limitations). As Levins has made for population biology, I will therefore describe the tension between the limitations protein scientists encountered, their aims for constructing models, and the complexity of proteinic objects. In so doing, I will discuss and analyze the strategy protein scientists worked out in order to manage this tension and to construct, then, what became, for them, a stable and efficient procedure of modeling.

With that interest in modeling practices, it seems then interesting to go back to Levins' work, which ties up model building and robustness analysis. In order to understand and analyze the procedure protein scientists have set up for constructing theoretical models of protein structure, as well as some properties of the models built, I will use Levins' analysis of modeling practices in population biology, which differentiates three main modeling strategies. But if Levins' analytical framework constitutes an interesting resource for discussing, in another scientific field, the type of modeling procedures used and the characteristics of the models constructed when using them, it seems also worthwhile to contrast the modeling strategy Levins uses and prefers, as a "population biology" modeler, and the modeling strategy worked out by protein scientists.

When contrasting these two approaches, it becomes notably clear that robustness analysis is a tool particularly well suited to Levins' modeling strategy. This is not the case with the modeling procedure protein scientists have set up within the strategy they chose. So, if the fruitfulness of the modeling strategy of Levins is partially related to idea that robustness analysis may provide a resource for determining which results are robust and, then, which models are trustworthy, the fruitfulness of protein scientists' modeling strategy and the stabilization of their modeling procedure have to be found elsewhere. One factor that seems have been important in the stabilization of this procedure of modeling is that the three limited resources

[3] See Wimsatt (2007), p. 6.

proteins scientists have used for devising the procedure have been mutually and iteratively adjusted. Such a mutual and iterative adjustment of theoretical, empirical and computational constraints has been implied in the stabilization of the modeling procedure, and, in so doing, in scientists' recognition that their procedure was efficient and their strategy fruitful.

So, my aim is to describe and analyze how, in a constrained epistemological context, protein scientists have used and mutually adjusted limited resources in order to construct what became, for them, an efficient procedure of modeling protein structure. As the efficiency of this procedure is partly related with its stabilization, I will pick out two factors implied in this stabilization. Thus, I will mention the mutual adjustments of the three constraints above-mentioned as well as the impact of the computational nature of the procedure on its evolving status, and on its diffusion within the scientific community. But if these two factors seem interesting for understanding to some extend how the modeling procedure has been stabilized and then conceived, by scientists, as being efficient, a real discussion of the trustworthiness of that procedure would require an analysis of the way the predictions produced with the models constructed have been tested against empirical data. Within the limited scope of this chapter, I will not discuss this point. I propose, then, a more modest approach, emphasizing some factors implied in the stabilization of the modeling procedure.

This chapter then discusses how practicing scientists, with limited resources, have exploited multiple constraints of construction and the limited tools computers are in order to construct and stabilize a modeling procedure. In so doing, I want to show that beside the justification of modeling results by robustness analysis and, more generally, the stabilization of a scientific procedure by a robustness scheme, others strategies of construction and stabilization of a scientific tool exist and have to be characterized. The strategy used by protein scientists to construct and stabilize their modeling procedure has been to manage multiple constraints of construction by mutually and iteratively adjusting, in a computational way, these theoretical, empirical and computational limited resources.

So as to understand more clearly the epistemological situation of modeling practices described by Levins for population biology, the nature of the tradeoffs he discussed, and the different strategies of model building he presented, I will, first, discuss Levins (1966) in Section 11.2. This discussion will be useful in order to specify, by using Levins' analytical framework, the type of modeling strategy used by protein scientists. Secondly, I will discuss, in Section 11.3, the specific constraints of the epistemological context within which theoretical approaches in protein chemistry have been developed. I will also specify, here, scientists' epistemic aims for constructing models of protein structure. Finally, I will analyze, in Section 11.4, the nature of the modeling procedure and of the models that protein scientists have constructed during the 1960s and 1970s. I will discuss, here, the theoretical, empirical and technological limited resources they could and had to take into account. To conclude, I will discuss, in Section 11.5, what is robustness for Levins and for Wimsatt in order to show, by contrast, the specificity of the strategy devised by protein scientists for constructing and stabilizing their theoretical procedure. I will emphasize, here, that the construction of the modeling procedure of proteins is based, and its stability depends, on a mutual and iterative adjustment of theoretical, empirical and technological constraints and on its special characteristics due to its computational nature.

11.2 Complex Systems, Brute Force Approach and Tradeoffs in Model Building

As Odenbaugh (2006) has argued, Richard Levins' article "The Strategy of Model Building in Population Biology" can be seen as a defense of a style of theorizing in population biology that Levins has developed during the 1960s with others biologists (Robert MacArthur, Richard Lewontin, E.O. Wilson, and others). His methodological essay is in particular composed of an analysis of three different strategies of modeling in the field of population biology. Levins argues that all these three strategies are relevant and important, and that it will not be beneficial to abolish this diversity in favor of one unique strategy. Levins' methodological discussion can then be seen as a defense of the richness of plural modeling strategies in population biology. As Odenbaugh (2006) writes: "(. . .) the target of Levins' 1966 article is ultimately *model monism* – that there is a single correct model type for successfully representing evolutionary-ecological systems" (p. 616, italics in the original). In defending this kind of "pragmatic pluralism" in theorizing,[4] Levins tries then to show that the modeling strategy he prefers and uses (which is one of the three strategies) is legitimate.

Levins considers that this pluralism is necessary for population biology (at least in the 1960s) because the way biological populations are described in his project (shared with others) of integrating population genetics and population ecology leads to have to "(. . .) deal simultaneously with genetic, physiological, and age heterogeneity within species of multi-species systems changing demographically and evolving under the fluctuating influences of other species in a heterogeneous environment" (Levins 1966, p. 18). As Weisberg (2006a) has noted, the core of Levins' article is constituted by the recognition of this complexity and by the two main options that are then discussed in order to deal with it.

Thus, Levins indicates that one way of modeling such a complex system would be "(. . .) to set up a mathematical model which is a faithful, one-to-one reflection of this complexity" (Levins 1966, p. 18). The aim of this first approach is to construct models that offer a representation as complete as possible of the complex system of interest. In this kind of models, every features of the system has to be represented within the model constructed. In order to reach this "ideal of completeness"[5] as far as possible when modeling the systems of population biology, one must then

[4] Wimsatt (2001) uses this expression, "pragmatic pluralism", when he discusses Levins' strategies of model building.

[5] This notion of an "ideal of completeness" is introduced and discussed by Weisberg (2006a, p. 626).

construct models constituted of a very large number of simultaneous partial differential equations with a lot of parameters. Once such a model has been constructed, it is then necessary to solve the equations to obtain numerical predictions, which have to be compared with known quantified properties of the system. Levins names this first way to deal with the complex systems of population biology "the brute force approach". But for him, such an approach is impracticable because too many parameters must be measured, because the equations are not analytically soluble and cannot be solved by the computers of that time, and because, even if solutions could be obtained, the results expressed in a long list of numbers would have no meaning for scientists.[6] Thus, Levins considers that this first approach cannot be followed, because of the complexity of the models *and* of the limitations of scientists.[7] Oppositely to the unreachable goal of this "brute force approach" – to construct a definitive model representing biological populations in their environment – a satisfactory theory in population biology is then, for him, what he calls "a cluster of models", that is a global coordination of different *idealized* models representing different sets of phenomena.[8] Levins' recognition of the impracticability of the "brute force approach" constitutes, thus, the core of his defense of model pluralism in population biology.

As Levins considers that this "brute force approach" cannot be followed, he then thinks that it is necessary to simplify the models constructed so as to work with manageable ones. This second approach is named by Weisberg (2006a) the "idealization approach". *It is within this second approach that Levins introduces a differentiation between three strategies of modeling*, which depends on the sacrifice made, concerning one of three desiderata of model building, when constructing simplified models. For Levins, besides the requirement of manageability of models, necessarily implied by the "idealization approach" chosen, the three desiderata of model building are generality, realism and precision.[9] Nevertheless, he considers that if it is desirable to maximize all these three aspects of models, it is not possible in practice; hence the tradeoffs in model building.[10] As these three aspects cannot be maximized if one wants to construct a manageable (then necessarily simplified) model, a sacrifice concerning one of these three aspects has to be made. Three strategies are thus possible: (1) "sacrifice generality to realism and precision"; (2) "sacrifice realism to generality and precision"; (3) "sacrifice precision to realism and generality" (Levins 1966, p. 19). Each of these strategies is then exemplified by the works of different groups of scientists interested in population biology. For the first one, Levins cites notably the works done in systems ecology. The second one is referred to the works

[6] For the precise formulations used by Levins, see Levins (1966, pp. 18–19.)

[7] Odenbaugh (2006) has convincingly defended this point.

[8] See Levins (1966, pp. 26–27), in particular Fig. 3 on p. 26.

[9] See the next paragraph for a discussion of the meaning of the terms "generality", "realism" and "precision".

[10] For a discussion of the logical versus in practice impossibility of such a maximization (i.e. of the logical versus in practice necessity of the tradeoffs), see Weisberg (2006a, p. 636).

of physicists-turned-ecologists. The third strategy is the one favored by Levins (and others). So, the three strategies identified set up something like a panorama of the different styles of theorizing represented in 1960s population biology.

The question of the meaning of the terms "precision", "realism" and "generality" used by Levins remains. In his chapter, he doesn't discuss precisely these terms and he uses these notions in a somewhat ambiguous way. Weisberg (2006a) has clearly picked out this situation and has then proposed an in-depth discussion of these notions. For him, generality corresponds to "(. . .) the number of target systems that a model can be applied to" (Weisberg 2006a, p. 634), realism to an assessment "(. . .) of how well the structure of the model represents the structure of the world or (. . .) an assessment of how close the output of the model matches some aspect of the target phenomenon" (Weisberg 2006a, p. 635), and precision corresponds to "(. . .) the fineness of specification of the parameters, variables, and other parts of model descriptions" (Weisberg 2006a, p. 636). So, the generality of a model is a property corresponding to the number of target systems this model describes or can describe. Concerning realism, it is a notion which corresponds to the accuracy of a model, understood (in Levins' uses of the term realism) *either* as the accuracy of the representation offered by the model *or* as the accuracy of the predictions of the model.[11] So as to distinguish these two meanings of the term realism, I will use, in what follows, "representational accuracy" (or "representational fidelity" – see the preceding footnote) and "predictive accuracy" (or "dynamical fidelity" – see the preceding footnote). Concerning now precision, it seems important to note that this notion is used in order to characterize model descriptions and not the outputs of models. Precision is then different from predictive accuracy because, in Levins' sense, precision doesn't characterize the output of a model.[12] It is also different from "representational accuracy" because precision concerns the fineness of specification (in particular of "quantitative specification") of the description used to model the system of interest, which is not equivalent to the quality of the representation of the system's structure offered by the structure of the model description (i.e. "representational accuracy").

It seems interesting to consider in some details the nature of the three modeling strategies in order to understand more precisely to what generality, realism and precision refer, and how these terms are used.[13] Models built within the first strategy maximize realism and precision and sacrifice generality. Here, we find

[11] Weisberg (2006a, p. 635) analyzes this ambiguity of the term realism in this way. See also Matthewson and Weisberg (2009, p. 181), who speak of "representational fidelity" of a model (which concerns "how well a model describes the causal structure of the target system") and of "dynamical fidelity" of a model (which concerns model's "predictions about the quantities of measurable attributes" of the target system).

[12] In this context, it will not make sense to speak of "predictive precision". It is then useless to speak of "representational precision", because precision is exclusively understood as "representational precision" (and not as "predictive precision").

[13] Here again, for an in-depth discussion of each of the three strategies, see Weisberg (2006a, pp. 637–640). I use Weisberg's discussion for my rough description of the three strategies.

the models of systems ecology, which are complex mathematical models describing in great details, with a lot of precisely specified quantitative parameters, very specific systems in order to provide accurate predictions. They lack generality because they are tailored to particular real systems, and they maximize realism (i.e. "predictive accuracy") and precision because they respectively produce accurate predictions, and describe with a high degree of specification the system of interest. Within this first strategy, computers are used to solve numerically the equations of the models.

Concerning now models built within the second strategy, they maximize generality and precision and sacrifice realism. Levins considers here that this is the strategy used by physicists-turned-ecologists, who construct models by "(. . .) setting up quite general equations from which precise results may be obtained" (Levins 1966, p. 19). These models lack realism (i.e. "representational accuracy") because the simple equations they use describe highly idealized systems (as it is the case, in physics, with frictionless systems or perfect gases, as Levins mentions). Nevertheless, these simplified systems are precisely specified (in particular in a quantitative way) within the models, which can, therefore, maximize precision. Furthermore, these precisely specified models could be applied to many systems, whence their generality. Within this strategy, it is sometimes possible to resolve analytically the equations. This can lead to compare the results then obtained with measured properties of the systems and to analyze precisely, as Weisberg (2006a) puts forward, the mathematical structure of the model used. It is then possible to try to construct less idealized models (i.e. more realistic ones).

Finally, with the third and last strategy, "very flexible models", as Levins calls them, are built. These models are often graphical rather than described by mathematical equation(s), and the parameters used in the model description are often only qualitative. The results produced are then qualitative and express often tendencies of evolution and contrasts between two situations in the form of inequalities. These models maximize generality because they are constructed in order to compare, for example, the effects of the same parameter on distinct phenomena. So, they have to be applicable to different analogous systems. Moreover, they maximize realism (i.e. "representational accuracy") because the aim of constructing this type of models is to test explanatory hypotheses concerning the ways a particular system evolves. It seems then necessary that such models characterize accurately these ways of evolving if we want the assumption to be explanatory. Nevertheless, as these models are deliberately qualitative ones, they don't offer a precise specification of the system being modeled, because the precise values for the parameters are not fixed. Thus, precision is sacrificed within this third strategy of model building, exemplified by Levins' works.

If the notions of "generality", "realism", and "precision" used by Levins remains perhaps somewhat ambiguous, they nevertheless enable to understand, at least intuitively, what are the characteristics of each strategy and the differences between all three. Moreover, by considering which specific tradeoff has been chosen when constructing a model, we can grasp the main goal assigned to model building by such or such scientists. For example, the first strategy is used when someone wants to make quantitative predictions about a very specific system, whereas the third strategy is used for trying to explain general features of phenomena.[14]

It is after having presented and discussed these three strategies that Levins introduces the notion of "robust theorem" and the idea of what we now call "robustness analysis". As robustness analysis in the sense of Levins will be discussed in conclusion, I will not develop at length, here, the way this notion is introduced. I just want to indicate what is a robust theorem and that robustness analysis fit in relatively well with the defense, by Levins, of model pragmatic pluralism. A robust theorem is not, for Levins, a theorem in the common sense of the term. It is more simply an explanatory hypothesis concerning the ways a particular biological system evolves. Levins takes for example the following statement to be a robust theorem: "in an uncertain environment species will evolve broad niches and tend toward polymorphism" (Levins 1966, p. 20). If several models, with different simplifications for describing the phenomenon but all incorporating the same biological hypothesis for trying to explain it, lead to similar results, then, in that case, these results don't depend on the particular simplifications of each model but on the biological explanatory hypothesis. This explanatory hypothesis, which consequences, when derived by various models with different simplifications, are similar, can then be considered as a robust theorem. As Levins puts it: "(. . .) if these models, despite their different assumptions, lead to similar results, we have what we can call a robust theorem that is relatively free of the details of the model. Hence, our truth is the intersection of independent lies" (p. 20).

So, the introduction of the notion of a robust theorem and of the underlying procedure of robustness analysis depends on the recognition that models are artificial constructions, which include artificial assumptions. It is because all models have artificial assumptions that constructing several alternative models of a same phenomenon becomes, for Levins, necessary, in order to try to secure a "theorem". Moreover, it has to be emphasized, here, that the recognition of this artificial character of models leads Levins to a defense of model pluralism, which is congruent with the way robustness analysis is conducted (by analyzing if a same result is obtained with *several alternatives models* using different assumptions and simplifications, in order to judge if this result can then be seen as trustworthy (i.e. robust)). So, robustness analysis is typically a tool linked with the model pragmatic pluralism defended by Levins, and this tool is particularly well suited to the third strategy of model building, exemplified by Levins' own works, because constructing different qualitative models of the same system is, here, relatively easy (it requires, nevertheless, that interesting assumptions are used). We could probably say, then, that the fruitfulness of this particular strategy of model building ("sacrifice precision to realism and generality" by constructing "very flexible models") is, in part, linked with the relatively easy possibility it offers to use robustness analysis.

In the two next sections, I will develop my case study on modeling practices in protein chemistry in the 1960s and 1970s. In so doing, I will try to show, as Levins

[14] In his presentation of the three strategies, Weisberg (2006a, pp. 637–640), discuss in more details the goals associated with each strategy.

has made for population biology, how the epistemological context, which articulates complexity of the modeled object *and* scientists' limitations, is important in order to understand the strategy protein scientists have devised for constructing theoretical models of proteins structure, as well as the status of the models built. I will examine some properties of these models, notably their generality and realism. Levins' essay constitutes therefore an interesting resource for discussing model-based activities in another scientific field.

11.3 Epistemological Situation of Theoretical Approaches in Protein Chemistry

In order to understand the epistemological situation of the theoretical approaches to proteins properties in the 1960s and 1970s, it seems necessary to present, schematically, what kind of molecular object proteins are. Proteins are organic compounds that play different fundamental functions within cells. These organic compounds are biological *macro*molecules, typically made of thousands of atoms. These biopolymers are composed of repeating structural units. Twenty such natural occurring structural units, named "amino acids", exist. Most proteins have the property of naturally folding into a precisely defined three-dimensional structure: scientists speak of the native *conformation* of a protein for this naturally occurring structure. A conformation of a molecule is thus a particular three-dimensional arrangement of its atoms. For one protein, different conformations are theoretically possible because of the various possible rotations around certain chemical bonds. The native conformation is then one among the great number of theoretically possible conformations of a protein: this collection of theoretically possible conformations of one protein is called its "conformational space", typically of an order of 3^{100} for a protein of 101 amino-acids.

The three-dimensional native structure of a protein is complex. It can be noted here that this structural complexity has immediately been recognized by the scientists (John Kendrew and Max Perutz) who, using X-rays scattering experiments, were able to propose in 1960 the first structures at atomic resolution of two proteins, namely myoglobin and hemoglobin.[15]

This molecular complexity of proteinic objects (huge number of atoms, intricacy of the folded structure, size of the conformational space) helps us understanding the epistemological situation of theoretical approaches in protein chemistry during the 1960s and 1970s. If, in order to grasp and deal with the complexity of the structures experimentally produced, some theoretical approaches were needed, and called for

[15] See Kendrew et al. (1960) and Perutz et al. (1960). Concerning the complexity of the structure, Kendrew wrote: "Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure" (Kendrew et al. 1958, p. 665). On the works of Perutz and Kendrew, see de Chadarevian (2002) and Debru (1983).

by scientists, the theory that *in principle* governs the properties of proteins, just as for any other molecular objects, was nevertheless not applicable *in practice* because of computational intractability. As shown by philosophers and historians interested in the question of the possible reduction of chemistry to physics, or more simply in quantum chemistry, the application of quantum mechanics to molecular systems has always been problematic and has led to increasingly complex and laborious computations.[16] That explains the central character of computers in the culture and practices of quantum chemistry after World War II. As quantum theory was already very difficult to apply to molecular systems of three, five or ten atoms, its use, even in conjunction with the specific theoretical descriptions and computational procedures developed between approximately 1930 and 1960 in quantum chemistry,[17] was clearly seen, by scientists, as definitively impracticable for proteins.

So, as for the situation in population biology described by Levins, there is a tension in protein chemistry between the complexity of the system under study *and* the limitations of scientists (the fact, here, that they are not computationally omnipotent). This tension has then led to the necessary development of a special modeling procedure, which doesn't use, at least directly, the theoretical formulations of quantum mechanics. Within this particular and constrained theoretical context, the impracticable approach (equivalent to the brute force approach criticized by Levins) is, more clearly than in the case of population biology, a (brute force) *theoretical application*. The modeling procedure devised by protein scientists is then an alternative approach, set up for allowing the construction of computationally manageable models.

Before the 1960s, a relatively long tradition of modeling structure and possible conformations of proteins already existed, but it was, within this tradition, *material* molecular models that were constructed (as for example by the famous chemist Linus Pauling).[18] Although material models were still used in the 1960s and 1970s, notably for representing the structures obtained by processing and interpreting X-rays experimental data,[19] practices of *theoretical* modeling also emerge during the 1960s. It is within such practices that the modeling procedure I am interested in has been developed.

As for all processes of emergence of a scientific practice, several factors can be put forward to understand this specific one. I will only mention here the scientists' epistemic aims that led to such an emergence. From the scientists' point of view, as noted above, the first need was the development of tools to analyze the great intricacy of the first structures experimentally obtained, and to test and refine the

[16] See in particular Scerri and McIntyre (1997), Schweber and Wächter (2000), and Park (2009, 2003).

[17] On these theoretical descriptions, which use various approximations, and these computational procedures developed in quantum chemistry, see for example Park (2009, 2003), Ramsey (2000, 1997) and Simoes (2003).

[18] See Francoeur (1997, 2001) for an historical analysis of material molecular models in chemistry (including protein chemistry).

[19] See de Chadarevian (2004).

structures that were constructed by processing and interpreting X-rays data: a very difficult task.[20] But, secondly, there was also a hope: if sufficiently good theoretical models of proteins could be devised on the basis of structural experimental data already obtained, then it would be possible, by exploring the conformational space of these macromolecules, to predict the native conformation of proteins – seen as the active one in cells – on the unique basis of a knowledge of their amino acids sequence. This would have potentially led to avoid the really laborious work to experimentally determine the three-dimensional structure of proteins. Moreover, this specific epistemic aim fitted in with the then current agenda of Molecular Biology.[21] Molecular Biology was interested, within the so-called "central dogma", in an understanding of genetic information flow from the one-dimensional structure of DNA (the sequence of bases) to the three-dimensional structure of proteins. The problem of predicting the 3-D structure of a protein from the knowledge of his sequence is known as the "protein folding problem",[22] a typically hot question for Molecular Biology in the 1960s and 1970s, and still today within structural genomics.

11.4 "Empirical Models" of Proteins: Status of the Procedure of Modeling and Resources for Its Construction

So far, we have seen what epistemic aims led scientists to construct theoretical models of protein conformations. But what resources could they exploit for such a construction? As noted above, since the use of theoretical formulations from quantum mechanics lead to non-manageable equations, even with the introduction of

[20] The construction of a three-dimensional structure of a molecule from X-rays data is a difficult work. Notably, the electronic density distribution of the molecule is calculated from the diffraction pattern, the electronic density distribution is then represented on electronic density maps, and a three-dimensional model of the molecule is constructed by using these maps. Thus, the structure proposed is the result of a complex analysis of X-rays data. When such a structure has been constructed, scientists try to test it, notably against stereochemical rules already adopted by the community. These tests lead to a refinement of the structure proposed. The modeling procedure I am interested in has been used in order to test and refine the structures proposed for various proteins. Others methods have also been used. For precisions concerning (the complexity of) X-rays data analysis, and structure refinement, see Perutz (1964), de Chadarevian (2002) Chapter 4, and de Chadarevian (2004).

[21] As Kendrew wrote in his Nobel lecture: "The geneticists now believe – though the point is not yet rigorously proved – that the hereditary material determines only the amino acid sequence of a protein, not its three-dimensional structure. That is to say, the polypeptide chain, once synthesized, should be capable of folding itself up without being provided with additional information; this capacity has, in fact, recently been demonstrated by Anfinsen *in vitro* for one protein, namely ribonuclease. If the postulate is true it follows that one should be able to predict the three dimensional structure of a protein from a knowledge of its amino acid sequence alone" (Kendrew 1964, pp. 676–98).

[22] The protein folding problem is called a problem because proteins have so many degrees of freedom; remember the size of the conformational space.

the approximations developed in quantum chemistry methods of ab initio or semi-empirical calculations, protein scientists have to find other theoretical resources. As the goal of constructing models of proteins was to gain knowledge of protein's conformations stability, protein scientists used a very simple theoretical formulation that has been proposed at the end of the 1940s,[23] and has been mainly used during the second half of the 1950s,[24] in order to understand the stereochemistry of organic compounds within the field of physical organic chemistry. It is not the place here to precisely discuss the origins, uses, transformations according to different contexts and the diffusion of this theoretical formulation.[25] This formulation stems notably from some attempts to interpret infrared spectra of (organic) molecules[26] and has been used, as already noted, in organic stereochemistry as well as in polymer chemistry. It seems more interesting, for the purpose of the present chapter, to write down this formulation in order to understand the characteristics exhibited by the models of molecules based on it.

The formulation defines a potential energy for a molecule for every set of positions of the atoms, that is for every conformation, as follows:

$$\mathrm{E} = \Sigma \left(u_0(r_0/r)^{12} - 2u_0(r_0/r)^6\right) + \Sigma ½ k_s(l - l_0)^2 + \Sigma ½ \mathrm{k_b}(\theta - \theta_0)^2$$

where l is a chemical bond distance, θ a bond angle, k_s and k_b are force constants, l_0 and θ_0 are equilibrium values of the bond length and angle, r is the distance between two interacting atoms, and $-u_0$ is the minimum value of the interaction energy (at $r = r_0$). The sum, for the first term, is made over all pairs of non-bonded atoms. For the second and third terms, the sums are made, respectively, over all pairs of bonded atoms and over all bond angles.

So, this simple formulation, at the heart of the models of protein that were constructed, involves a particular representation of matter: molecules are constituted of valence-bonded 'atoms' (and not of nucleus and electrons as in quantum mechanics), and are roughly speaking represented by a system of balls connected by springs. This particular idealization shows that the question of the very accuracy of the representation offered by that type of protein models is not a priority for scientists. They obviously know that this representation is not accurate, but they adopt it precisely because it is useful, because it is the unique representation at hand that can lead to computationally manageable models, but, also, because it is a representation which is consistent with a classical conception of molecules, as conveyed, for example, by material molecular models and by some analysis of molecular vibrations. For scientists, the models constructed on the basis of this formulation use a relatively usual idealization in chemistry. The representation they offer is then acceptable but not

[23] See Hill (1946) and Westheimer (1947).

[24] See Westheimer (1956).

[25] For details, see Wieber (2005), Chapters 6 and 7.

[26] See for example the pioneering works of physical chemist Bjerrum, as described by Assmus (1992), and the subsequent works of Wilson et al. (1955).

accurate. Thus, the validity of the models can only be above all pragmatic: it will only be possible, for scientists, to test their validity by using these models and by comparing, then, the predictions obtained with known and accepted empirical data.

We can then see, here, that the modeling strategy devised by protein scientists sacrifices the realism of the representation of molecules (i.e. the "representational accuracy" of models) to computational imperatives. But scientists hope that if sufficiently good parameters were used, the predictions obtained would be reasonably accurate. If the models constructed on the basis of this simple theoretical formulation are not realists, in the sense of "representational accuracy", they could nevertheless "numerically describe", with sufficient accuracy, the structural properties of proteins. These models of protein structure could then be viewed as being realists in the sense of "predictive accuracy". The double meaning of Levins' "realism" ("representational accuracy" *or* "predictive accuracy") is here particularly manifest, because, in this context, a well-confirmed theory, with a precise ontology, governs the systems of interest but cannot be applied to these systems. So, because the manageable models that can be constructed use another ontology, they lack "representational accuracy". But scientists hope that these models could nevertheless offer accurate predictions, that they will have a good "predictive accuracy".

There is a second interesting and, for scientists, fundamental characteristic of the simple theoretical formulation used within the modeling strategy. In order to construct a model of a particular protein (the theoretical formulation is obviously not, by itself, a model of protein), one has inevitably to fix the values of parameters appearing in the theoretical formulation, for all pairs of non-bonded atoms, for all pairs of bonded atoms and for all bond angles. And the number of parameters is really important, because a protein is made of different types of atoms and of chemical bonds. As the parameters used are *empirical parameters*, we have thus to note here, firstly, that this modeling procedure is called by scientists "empirical modeling", and, secondly and more importantly, that the strategy of using this modeling procedure is very dependent on the availability of the empirical data required. We recognize, here, the problem of the measurement of a great number of parameters that Levins has stressed when discussing the brute force approach in population biology.

For protein models, as well as for models of other organic molecules, different types of empirical data are needed: infrared spectroscopic data, crystallographic data, thermodynamic data *etc.*. . . Thus, scientists who want to construct a model for a particular molecule must find, one the one hand, what data are available for that molecule, and choose, and the other hand, which values of data it seems preferable to use when different values are available for one type of data. Of course, all the data needed are never at hand for the particular molecule of interest, and they are then estimated and adjusted, by some kind of theoretical tinkering, to the specific case of that particular molecule of interest, from the data available for other molecules. It is important here to stress that such modeling practice of molecular objects couldn't have been developed without the revolution of physical instrumentation in chemistry since the 1930s.[27] But it is equally necessary to remind here the complexity of proteinic objects. Since these objects are constituted of a huge number of atoms, the use of physical instrumentation to obtain typical data for these molecules was very difficult; hence the amount of data necessary to parameterize a model of protein was really thin. The work of estimating and adjusting empirical data was thus more extensive in protein chemistry than in organic chemistry, where more data were available, because smaller molecules are studied. To conclude this point, we can stress that for constructing models of molecules within this modeling strategy, scientists had to exploit creatively, within a practice of theoretical tinkering, some empirical resources. The parameterization is the central stage in the procedure of modeling, and it demands a good knowledge of empirical results for such or such type of molecule, and specific skills to make and justify the choices and adjustments of data.[28] Finally, different research teams made these choices locally, and different sets of parameters have been constructed during the 1960s, in organic chemistry as well as in protein chemistry.[29] Scientists speak of a "force field" for a set of parameters and equations, because a parameterized equation describes the potential energy for a molecule.

I turn now to the third resource that protein scientists used when constructing their "empirical models". This third resource is a technological one, namely computers. So, if a protein scientist has made the choice of using the theoretical formulation we have seen above, and has constructed a model for a molecule by choosing, estimating, adapting, adjusting different types of empirical data, he can now use this model to study the stability of some conformations or to refine (by minimizing the potential energy of the molecule) the structure proposed when interpreting X-rays patterns. But to do all that, it is necessary to calculate the potential energy of one or several conformations. When the modeling strategy was used in organic chemistry in the 1950s without the help of computers, the task of calculating all the chemical bonds geometries and energies and all the interactions between all pairs of non-bonded atoms was still complex and really laborious. But a pencil and paper application of the method to bigger organic molecules and a fortiori to proteins was

[27] On the transformations in chemistry induced by the spreading of physical instrumentation, see Morris and Travis (2003).

[28] Choosing and adjusting empirical data in order to construct a set of parameters demands a good knowledge of empirical results in chemistry, a good appreciation of the validity of such or such empirical technique for measuring such or such property of such and such molecule, as well as analogical reasoning and extrapolations in order to decide, for example, how to construct a particular parameter concerning an interaction between two "atoms" within a particular molecule from the empirical value of that interaction between this two same (or chemically similar) "atoms" in another molecule.

[29] For details about the situation at the end of the 1960s, and references, see the review of Williams et al. (1968) for organic chemistry, and the review of Scheraga (1968) for protein chemistry. During the 1970s, sets of parameters continued to be developed and refined. For references, see Wieber (2005), Chapters 6 and 7.

out of reach.[30] The development and spread out of that type of modeling practices in protein chemistry (and more generally in chemistry) has thus been fully dependent of the use of computers. These practices would not have been efficient if these technological instruments of computation had not been available.

But if computers were needed for that efficiency, the use of these calculating machines altered modeling practices in turn. So, to define precisely the characteristics of each atom inside a protein according to their molecular surrounding, scientists were able to use an increasingly large number of parameters stored in databanks, which the computer program could access quickly. A mode of calculation based on pencil and paper would not have allowed such increase in the number of parameters used for modeling, because it would not have been manageable. And with this greater number of parameters quickly accessible, models of more and more proteins could then be conveniently constructed and used. Finally, the computational nature of the modeling practices has also allowed a type of crystallization and spreading of the choices made locally concerning the empirical parameters, and more generally a stabilization of the modeling procedure, thanks to the construction and dissemination of computer programs packages.[31]

This partial black-boxing as computer software of this procedure seems fundamental in order to understand its increasing stability. With the construction and dissemination of these computer programs packages integrating the procedure of modeling, the community of its users has been broadened. In this process, the theoretical tools constructed have been integrated, thanks to their computational nature, to the classical toolbox used by experimenters for processing and interpreting empirical data of molecular structure. The procedure of modeling has then participated to the production of more and more experimental results. In this sense, many experimental results depend, today, on this procedure. Following Wimsatt (2007), it seems then possible to consider that the procedure of modeling has gained in stability by being "generatively entrenched".[32]

We can now conclude on this strategy of modeling in protein chemistry by discussing the status of the procedure of modeling, the properties of the models devised within this strategy, and the fundamental character of the *computational* nature of these models. As we have seen, these models don't offer "representational accuracy" because they are not constructed by applying the theory that governs protein properties. As an application of this theory is not possible in practice, scientists sacrifice

[30] Hendrickson (1961) constitutes the first *computational* application of this procedure of modeling in organic chemistry. For proteins, the development of this modeling procedure has always been computational; see for example Scott and Scheraga (1966).

[31] For modeling proteins (and more generally biological macromolecules and even organic molecules), three main packages were developed during the 1970s, by three different research teams: see Momany et al. (1975), Weiner and Kollman (1981), and Brooks et al. (1983). Packages specifically dedicated to modeling organic molecules were also developed during the 1970s.

[32] Wimsatt (2007) defines "generative entrenchment" in this way: "A deeply generatively entrenched feature of a structure is one that has many other things depending on it because it has played a role in generating them" (pp. 133–134).

deliberately this type of accuracy by choosing another theoretical formulation, which is applicable to the systems of interest and could conduct to relatively accurate predictions. If this choice of theoretical formulation impacts the status of the models constructed by defining what kind of realism they will hold, it has equally interesting consequences concerning the generality of the models and of the procedure of modeling. Thus, this choice of theoretical formulation leads to a lot of empirical parameters. But the parameters used with success for modeling one particular molecule cannot be used, strictly speaking, to construct a model for another molecule: each parameterized term in the formulation has no real meaning in itself, and only global numerical results obtained when applying the whole parameterized formulation can have a real meaning if they are accurate, that is if they are considered as good predictions. So, the parameters chosen, estimated and adjusted for one protein, within a practice of theoretical tinkering, are theoretically not transferable for another one. In that respect, the models constructed within this procedure of modeling greatly lack generality. But this is not the whole story. As the procedure would not be useful if it was necessary to reconstruct, each time, for each new protein, the parameters, a hypothesis of transferability is made.[33] With such hypothesis, a gain in generality is obtained, not for each model constructed, but for the procedure of modeling. However, the in practice transferability of each parameters has to be shown pragmatically by using these parameters for constructing more and more models of different proteins and by testing against empirical results the outputs obtained with these models. The computerization of the procedure of modeling is then really fundamental: with more and more models constructed and effective calculations executed, scientists have been able to increasingly test the results produced against empirical data in order to *iteratively optimize* the parameters chosen for modeling. Moreover, a large number of different parameters, suitable for more and more types of molecules, have been stored in computer programs, as indicated above. So, the use of computers has allowed an increase in generality of the "force fields" elaborated, and the construction of models that scientists consider as more trustworthy. Nevertheless, the procedure of modeling is such that a "force field" is only validated by its usage, by the accurate predictions obtained for circumscribed families of molecules, and its generality is then inevitably limited. Scientists are then lead to perpetually refine the parameters stored in computer programs, and the choice of a particular "force field" depends on the type of molecule studied and on the question asked concerning this molecule.

11.5 Conclusion

As noted in introduction as well as in Section 11.2, robustness analysis emerges, in Levins' works, within a discussion concerning the constraints set on modeling practices by the specific epistemological context in population biology. These constraints

[33] On the question of the transferability of parameters, see Burkert and Allinger (1982, pp. 3–4).

have led to the development of three different modeling strategies, implying different tradeoffs. It is because Levins (1966) considers that all models have artificial assumptions, and that "there is always room to doubt whether a result depends on the essential of a model or on the details of simplifying assumptions" (p. 20), that a method becomes then necessary in order to judge the trustworthiness of a particular result obtained when using one particular model. Robustness analysis is precisely, for Levins, such a method: if a same result is obtained with several alternatives models using different assumptions and simplifications, this result can then be seen as trustworthy (i.e. robust). Thus, robustness analysis has to be understood as the final stage of an epistemological strategy where multiple alternative models of a same system are deliberately, and initially, constructed. Robustness analysis is based on the examination of the results produced when using these multiple alternative models, and constitutes the core of the strategy. We can speak of a heuristic procedure for this "robustness strategy", which is developed in order to take into account that all models have artificial assumptions.

This characteristic of models follows from the above-mentioned specific constraints of the epistemological context in population biology. It is because scientists face practical limitations with respect to what they can compute, measure and understand when modeling the complex systems of population biology that the models they construct entail simplifications and artificial assumptions, whose consequences are managed by using the heuristic strategy which includes robustness analysis. And if such a strategy is used, and a result said to be robust, we are then more confident with respect to the trustworthiness of the models that have led to this particular result. As Weisberg (2006b) considers, this method is then useful "(. . .) for determining which models make trustworthy predictions and which models can reliably be used in explanations" (p. 731).[34]

Wimsatt has expanded the notion of robustness to a broader set of procedures and contexts of use. When he speaks of robustness, a general scheme concerning the constitution of the solidity[35] of an entity, or a property, or a relation, or a proposition is used. As Soler points out in Chapter 1, Wimsatt's notion of robustness "(. . .) refers to the idea of the invariance of a result under multiple independent determinations". For Wimsatt, this scheme is conceived as being very general, and it can be used in different contexts in order to distinguish "(. . .) that which is regarded as ontologically and epistemologically trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting" (Wimsatt 1981, p. 128). Wimsatt considers that "a family of criteria and procedures" (Wimsatt 1981, p. 126), based on

[34] It seems interesting to note, here, that Weisberg (2006b) considers robustness analysis to be an important method in sciences where complex systems "(. . .) have yet to be described by comprehensive theories" (p. 731). When we construct models of a system which is governed by a well-developed theory, this theory "(. . .) could be used to determine how much distortion was introduced by each idealization [in each model]" (p. 731). As he puts it: "(. . .) theories have the resources to estimate the effect of various idealizations, providing guidance about what must be included when particular degrees of accuracy and precision are required" (p. 731).

[35] I use, here, the term "solidity", as introduced by Soler in Chapter 1.

this robustness scheme, are used to this end. He mentions a long list of such procedures, for example: "(a) using different sensory modalities to detect the same property or entity (. . .); (b) using different experimental procedures to verify the same empirical relationships or generate the same phenomenon [. . .]; (c) using different assumptions, models, or axiomatizations to derive the same result or theorem (. . .)" (Wimsatt 1981, pp. 126–127). All these procedures can be called, for Wimsatt, "robustness analysis". Thus, when someone uses such a procedure, he or she analyzes what has been obtained with different (at least partially) independent derivations in order to establish if a robustness scheme could be find. If such a scheme is found, Wimsatt considers that we have more reasons, then, to judge the result as being reliable.

So, Wimsatt has proposed a fruitful generalization of the notion of robustness with respect to its more restricted sense in Levins' works. By delimiting a general robustness scheme, robustness analysis can then be extended to others procedures – notably to the "triangulation" of a same result by different empirical procedures – which was not discussed in the very specific context of modeling in population biology. Moreover, robustness analysis is then not only a procedure used by practicing scientists in order to judge the trustworthiness of their results and models, but it becomes also, for philosophers of science, an explicit scheme for describing the way scientists try to secure the results they produce and a good starting point for discussing scientists' judgments of robustness and possible bias in the methodologies and reasoning they have used for producing such or such result. Finally, if the notion of robustness is more general for Wimsatt, it seems equally that he uses the term "robustness" with multiple senses, exploiting then fully the resources offered by the most general form of that notion.[36]

We have seen that Levins' robustness analysis really makes sense within a strategy, which is developed because models have limitations imposed by a constrained epistemological context in biological modeling. Here again, Wimsatt generalizes this idea of robustness as an epistemological strategy, in the context of a conception of science within which the limited capacities of practicing scientists are fully recognized. Thus, the importance of robustness (and more generally of heuristic procedures) is linked, for Wimsatt, with the necessary recognition that the world is complex, that practicing scientists have limited capabilities and are fallible, that they are not omniscient and computationally omnipotent, and that the tools they

[36] In his recent review of Wimsatt (2007), Calcott (2011) distinguishes three kinds of robustness, each occurring in Wimsatt's book: robust theorems (in the sense of "robustness" conveyed by Levins' robustness analysis), robust detection (or triangulation, that is to say, the production of the same result by different and at least partially independent empirical procedures), and robust phenomena (in the sense that a system, a mechanism, is robust when it "continues to function reliably, despite perturbations or interventions", as Calcott wrote). So, the sense of "robustness" used by Wimsatt presents some kind of multiplicity, but behind that multiplicity, there is probably a common structure to all these kinds of robustness, as Calcott suggests and discusses.

use have limitations.[37] When we recognize this overall situation, when we resist "in principle" claims about the way science works, then robustness analysis really makes sense and becomes the core of an important strategy (among others).

The case study on modeling practices in protein chemistry I have presented and discussed is definitively linked with this picture of science interested by the limitations scientists encounter and the "in practice" strategies they have then to devise in order to study complex systems. Nevertheless, the strategy used by protein scientists to construct and stabilize their modeling procedure cannot be described by a robustness scheme. But before pointing out the particular scheme of construction and stabilization that emerges from my case study, it seems interesting to go back to Levins' analytical framework of modeling practices.

This framework, devised for analyzing different styles of theorizing in population biology, constitutes an interesting resource for discussing the modeling procedure developed in protein chemistry during the 1960s and 1970s and the characteristics of the models constructed when using it, by examining their realism, precision, and generality. As we have seen in Section 11.4, these models lack generality (even if the hypothesis of transferability of parameters and the use of computers have allowed an increase in generality of the *procedure* of modeling) as well as they lack realism (in the sense of "representational accuracy"). But scientists try to construct and use realist (in the sense of "predictive accuracy") and precise models (because the representation used to model the system of interest is precisely specified, with quantitative parameters). So, the strategy devised by protein scientists seems close to the first model building strategy discussed by Levins ("sacrifice generality to realism and precision"). Moreover, as it has been emphasized, the use of computers was indispensable for developing and using the protein modeling procedure set up within the chosen strategy, as it was also the case for the models of systems ecology, which exemplify, for Levins, the models constructed within the first strategy he discussed. Levins' analytical framework is then an interesting tool for characterizing the modeling strategy used by protein scientists. And such characterization is finally interesting because it leads us to contrast this strategy with the one defended by Levins, and to ask what made these two strategies fruitful and efficient strategies.

Robustness analysis is typically an argument showing the fruitfulness and efficiency of the type of qualitative modeling strategy Levins practiced, because it is a tool particularly well suited to this modeling strategy: with "very flexible models", multiple models of a same system can be constructed in a relatively easy way. Concerning the strategy devised by protein scientists, its fruitfulness is not associated with robustness analysis but with its stabilization, which cannot be described by using a general robustness scheme.

[37] See, for example, Wimsatt (2007), in particular the introduction and the epilogue, or Wimsatt (1981, pp. 151–153). It seems worth noting here that all scientific tools have, for Wimsatt, limitations. So, models have limitations (as Levins points out), as it is equally the case for our sensory modalities, our measurement apparatuses, etc.... Within such a generalized constrained epistemological context, the importance of "robustness strategy" is also generalized.

Here, we have to recognize how the *computational* nature of the procedure of modeling devised in protein chemistry has been fundamental in its stabilization, which has participated to the recognition, by scientists, that this procedure was efficient and fruitful. So, as I have tried to show, its computerization has allowed the construction of more and more models and the effectuation of more and more effective calculations. Scientists have then been able to increasingly test the results produced against empirical data in order to iteratively optimize the parameters chosen for modeling. With such a computational optimization of parameters, and with the large numbers of parameters stored in computer databanks and quickly accessible by the computer program, the procedure of modeling has gained a (limited) *generality*, which seems have been important in its stabilization and in the widest recognition of its relevance. After all, it is possible to consider that this procedure has acquired stability within a process of mutual and iterative adjustment of theoretical, empirical and computational constraints, because the three kinds of limited resource protein scientists have used for devising the procedure of modeling were mutually dependent, and limited: computational limitations has led to the adoption of a theoretical formulation which requires to fix, within a practice of theoretical tinkering (necessary because the amount of data necessary to parameterize a model of protein is thin), the values of a lot of empirical parameters which are then computationally optimized... But if an adjustment of these different constraints has been fundamental for the constitution of a stable procedure of modeling protein structural properties, the partial black-boxing as computer software of this procedure has also, as we have seen, increased its stabilization: the dissemination of computer programs packages integrating this procedure has broadened the community of its users.

Here again, we can emphasize that the computational nature of these modeling tools is obviously important. More generally, the technological characteristics of computers – the way they function, the limitations of their processing power, their accessibility for scientists – have a great impact on the evolving epistemic status of that kind of modeling practices, which have led to the emergence of what has been considered, by scientists, as a theoretical knowledge about proteins structure and stability.

Acknowledgements Many thanks to Léna Soler, Catherine Dufour, and Emiliano Trizio for helpful comments on previous drafts.

References

Assmus, A. 1992. "The Molecular Tradition in Early Quantum Theory." *Historical Studies in the Physical and Biological Sciences* 22:209–31.

Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. 1983. "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." *Journal of Computational Chemistry* 4:187–217.

Burkert, U., and N.L. Allinger. 1982. *Molecular Mechanics*. Washington, DC: American Chemical Society.

Calcott, B. 2011. "Wimsatt and the Robustness Family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings." *Biology and Philosophy* 26:281–93.

Debru, C. 1983. *L'Esprit des Protéines*. Paris: Hermann.

De Chadarevian, S. 2002. *Designs for Life. Molecular Biology After World War II*. Cambridge: Cambridge University Press.

De Chadarevian, S. 2004. "Models and the Making of Molecular Biology." In *Models. The Third Dimension of Science*, edited by S. de Chadarevian and N. Hopwood, 339–68. Stanford: Stanford University Press.

Francoeur, E. 1997. "The Forgotten Tool: The Design and Use of Molecular Models." *Social Studies of Science* 27:7–40.

Francoeur, E. 2001. "Molecular Models and the Articulation of Structural Constraints in Chemistry." In *Tools and Modes of Representation in the Laboratory Sciences (Boston Studies in the Philosophy of Science, Volume 222)*, edited by U. Klein, 95–116. Dordrecht: Kluwer.

Hendrickson, J.B. 1961. "Molecular Geometry. I. Machine Computation of the Common Rings." *Journal of the American Chemical Society* 83:4537–47.

Hill, T.L. 1946. "On Steric Effects." *Journal of Chemical Physics* 14:465.

Kendrew, J.C. 1964. "Myoglobin and the Structure of Proteins. Nobel Lecture, December 11, 1962." In *Nobel Lectures, Chemistry 1942–1962*, 676–98. Amsterdam: Elsevier.

Kendrew, J.C., G. Bodo, H.M. Dintzis, R.G. Parrish, H.W. Wyckoff, and D.C. Phihips. 1958. "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis." *Nature* 181:662–6.

Kendrew, J.C., R.E. Dickerson, B.E. Strandberg, R.G. Hart, D.R. Davies, D.C. Phillips, and V.C. Shore. 1960. "Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å resolution." *Nature* 185:422–7.

Levins, R. 1966. "The Strategy of Model Building in Population Biology." Reprinted in *Conceptual Issues in Evolutionary Biology*, 1st ed. (1984), edited by E. Sober, 18–27. Cambridge: MIT Press.

Levins, R. 1968. *Evolution in Changing Environments*. Princeton, NJ: Princeton University Press.

Matthewson, J., and M. Weisberg. 2009. "The Structure of Tradeoffs in Model Building." *Synthese* 170:169–90.

Momany, F.A., R.F. McGuire, A.W. Burgess, and H.A. Scheraga. 1975. "Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids." *Journal of Physical Chemistry* 79:2361–81.

Morris, P.J.T., and A.S. Travis. 2003. "The Role of Physical Instrumentation in Structural Organic Chemistry." In *Companion to Science in the Twentieth Century*, edited by J. Krige and D. Pestre, 715–40. Londres: Routledge.

Odenbaugh, J. 2006. "The Strategy of 'The Strategy of Model Building in Population Biology'." *Biology and Philosophy* 21:607–21.

Park, B.S. 2003. "The 'Hyperbola of Quantum Chemistry': The Changing Practice and Identity of a Scientific Discipline in the Early Years of Electronic Digital Computers, 1945–65." *Annals of Science* 60:219–47.

Park, B.S. 2009. "Between Accuracy and Manageability: Computational Imperatives in Quantum Chemistry." *Historical Studies in the Natural Sciences* 39:32–62.

Perutz, M. 1964. "X-Ray Analysis of Haemoglobin. Nobel Lecture, December 11, 1962." In *Nobel Lectures, Chemistry 1942–1962*, 653–73. Amsterdam: Elsevier.

Perutz, M.F., M.G. Rossmann, A.F. Cullis, H. Muirhead, G. Will, and A.C.T. North. 1960. "Structure of Haemoglobin. A Three-Dimensional Fourier Synthesis at 5,5 Å Resolution, Obtained by X-Ray Analysis." *Nature* 185:416–22.

Ramsey, J.L. 1997. "Between the Fundamental and the Phenomenological: The Challenge of 'Semi-Empirical' Methods." *Philosophy of Science* 64:627–53.

Ramsey, J.L. 2000. "Of Parameters and Principles: Producing Theory in Twentieth Century Physics and Chemistry." *Studies in History and Philosophy of Modern Physics* 31:549–67.

Scerri, E.R., and L. McIntyre. 1997. "The Case for the Philosophy of Chemistry." *Synthese* 111:213–32.

Scheraga, H.A. 1968. “Calculations of Conformations of Polypeptides.” *Advances in Physical Organic Chemistry* 6:103–84.
Schweber, S., and M. Wächter. 2000. “Complex Systems, Modelling and Simulation.” *Studies in History and Philosophy of Modern Physics* 31:583–609.
Scott, R.A., and H.A. Scheraga. 1966. “Conformational Analysis of Macromolecules. III. Helical Structures of Polyglycine and Poly-L-Alanine.” *Journal of Chemical Physics* 45:2091–101.
Simoes, A. 2003. “Chemical Physics and Quantum Chemistry in the Twentieth Century.” In *The Cambridge History of Science, vol. V. The Modern Physical and Mathematical Sciences*, edited by M.J. Nye, 394–412. Cambridge: Cambridge University Press.
Weiner, P.K., and P.A. Kollman. 1981. “AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions.” *Journal of Computational Chemistry* 2:287–303.
Weisberg, M. 2006a. “Forty Years of ‘The Strategy’: Levins on Model Building and Idealization.” *Biology and Philosophy* 21:623–45.
Weisberg, M. 2006b. “Robustness Analysis.” *Philosophy of Science* 73:730–42.
Westheimer, F.H. 1947. “A Calculation of the Energy of Activation for the Racemization of 2,2′-Dibromo-4,4′-Dicarboxydiphenyl.” *Journal of Chemical Physics* 15:252–60.
Westheimer, F.H. 1956. “Calculation of the Magnitude of Steric Effects.” In *Steric Effects in Organic Chemistry*, edited by M.S. Newman, 523–55. New York: Wiley.
Wieber, F. 2005. “Construction et Applicabilité Théoriques en Chimie des Proteins (1960–1990). Essai Epistémologique et Historique.” PhD thesis, Paris: Paris Diderot University.
Williams, J.E., P.J. Stang, and P. von R. Schleyer. 1968. “Physical Organic Chemistry: Quantitative Conformational Analysis; Calculation Methods.” *Annual Review of Physical Chemistry* 19:531–58.
Wilson, Jr. E.B., J.C. Decius, and P.C. Cross. 1955. *Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra*. New York: McGraw-Hill.
Wimsatt, W.C. 1981. “Robustness, Reliability, and Overdetermination.” In *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins, 124–63. San Francisco, CA: Jossey-Bass.
Wimsatt, W.C. 1994. “The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets.” In *Biology and Society: Reflections on Methodology, Canadian Journal of Philosophy*, Supplementary Vol. 20, edited by M. Matthen and R. Ware, 207–74.
Wimsatt, W.C. 2001. “Richard Levins as Philosophical Revolutionary.” *Biology and Philosophy* 16:103–8.
Wimsatt, W.C. 2007. *Re-Engineering Philosophy for Limited Beings. Piecewise Approximations to Reality*. Cambridge: Harvard University Press.

Chapter 12
Understanding Scientific Practices: The Role of Robustness Notions

Mieke Boon

12.1 Introduction

The rationale for considering robustness as an important notion for understanding scientific practices is directly related to key issues in the philosophy of science such as: why science is so successful; why theories are accepted; how scientific knowledge can be justified; and how scientific theories are related to the real world. Within this scope, which focuses on theories as the aim of science, the issue is whether robustness functions as a 'truth-maker' of theories or as an alternative to their truth.

The perspective from which robustness will be considered in this article is for understanding scientific research in the context of practical applications, such as technological, (bio-) medical and agricultural research, and the forecasting of natural processes. Scientific research in these fields interprets practical problems or technological functions in terms of *phenomena* that determine the cause of technological (dys)functioning. Scientific research aims at intervening with these phenomena (e.g. their artificial production or prevention) by developing scientific understanding about them (cf. Boon 2006, 2009; Boon and Knuuttila 2009). Therefore, the epistemic aim of these practices differs from the ultimate aim that the philosophy of science usually ascribes to science. The epistemic aim of scientific research in the context of such things as practical applications is the reliability and relevance of theoretical knowledge regarding these applications, rather than the truth of theories. From this practice-oriented perspective, accounting for the *truth of conclusions* drawn from scientific theories is more important than accounting for the *truth of scientific theories*. Clearly, someone may object that the distinction

M. Boon (✉)
Department of Philosophy, University of Twente, Enschede, The Netherlands
e-mail: M.Boon@utwente.nl

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_12,

between true theories and true conclusions is already accounted for in the relationship between fundamental and applied sciences, according to which drawing conclusions from true knowledge produced in the fundamental sciences produces true conclusions in the applied sciences. However, it is an empirical fact of scientific practices that the conclusions drawn from supposedly (approximately) true theories are often false (cf. Cartwright 1983) or irrelevant. Usually a significant amount of additional experimental research is required to produce scientific models that fit the real world (see also Morrison and Morgan 1999).

Given this situation, scientific practices in the context of practical applications require a philosophical account that explains what science can and cannot do, rather than an account of how the truth of scientific results is justified. Therefore, understanding the character and justification of the reliability and relevance of scientific knowledge used in practical applications is an issue that matters. Within this context, the principle question of this article is: *What part can robustness notions play in understanding scientific practices aimed at producing reliable and relevant scientific results (such as scientific theories, models and concepts, but also phenomena and physical systems) for practical application*? My thesis comprises four statements: (a) different kinds of robustness notions must be distinguished (i.e. metaphysical, regulative, methodological, ontological and epistemological), each pointing at different aspects and presuppositions of scientific research; (b) they have to be viewed as complementary to each other; (c) they are ultimately held together by the regulative principle 'same conditions – same effects'; and (d) robustness as an epistemological notion functions as an alternative to truth.

The structure of this article is as follows. Section 12.2 explains some philosophical presuppositions about scientific practices that provide the foundation for my argument. Section 12.3 presents a conceptual analysis of robustness. There appear to be different kinds of robustness notions. By utilizing traditional philosophical accounts of the justification of scientific knowledge, it also analyses how robustness is related to truth. This analysis seeks to examine whether robustness can be an alternative to truth. Section 12.4 explains why methodological robustness notions justify the attribution of epistemological (and ontological) robustness notions to scientific results, and why the role of 'same conditions – same effects' as a regulative principle is crucial.

My argument is divided into two parts, each of which takes a different approach. The first part (Section 12.3) uses traditional analytical approaches in the philosophy of science. It employs Van Fraassen's (1980) notion of empirical adequacy as a philosophical guide for articulating the role of epistemological criteria in accepting epistemological results. The second part (Section 12.4) focuses on the idea that scientific practices seek a variety of scientific results. It takes Hacking's (1992, 1999) notion of a mutual fit between different elements that constitute a laboratory practice ('ideas, matériel, and marks') as a preliminary philosophical account in which 'true theories' are no longer regarded as crucial for explaining the success of science.

12.2 Traditional Philosophy of Science Versus Philosophy of Scientific Practices

My approach in Part I (Section 12.3) of the argument aims to tie in with a traditional approach in the philosophy of science because this tradition contributes to the conceptual clarification of robustness. At the same time, traditional philosophical approaches involve several assumptions about science that are unproductive for understanding concrete scientific practices. In this section, I will compare some of the important presuppositions of traditional philosophy of science with those I consider as more appropriate for a philosophy of scientific practices. The proposed alternatives will be taken as the philosophical foundation to Part II (Section 12.4) of my argument.

As an alternative to the assumption that the aim of science is true theories, I propose that the epistemic aim of science is to produce epistemic means that allow for scientific reasoning about the (natural or laboratory) world (see also Rouse 2009, 2011), and as a consequence, that the task of the philosophy of science is to account for the acceptance of scientific results that facilitate the performance of this epistemological function. This alternative assumption does not necessarily exclude the role that truth could play. Rather, this proposal is made because accounting for the truth of theories is extremely problematic and may not even be necessary in accounting for the success of science and understanding actual scientific practices.

The second assumption of traditional approaches is the idea that science can be reduced to two basic elements: facts and theoretical knowledge. Observations and data are considered as the objective basis of facts, meaning that facts are philosophically unproblematic. The role that facts are supposed to play is in proving theories. The divide between the two elements is crucial to accounts of methodologies that justify (or falsify) the truth of theoretical knowledge, such as induction or verification (confirmation or falsification) by hypothetical-deductive approaches.

The so-called ‘New experimentalists’[1] have criticized the idea that facts result from observations and data in an unproblematic manner. They have emphasized that observations and data of the independent real world are gathered by means of experiments, technological instruments and data-processing. Therefore, facts result from constructive activities in the physical world, while these constructive activities go together with practical and theoretical reasoning about technological instruments, data and physical phenomena. What is more, data, facts (i.e. descriptions of observable physical phenomena), data-processing, experiments, instruments and theoretical interpretations develop in a mutual interplay, and eventually ‘vindicate’ one another. Hacking (1992), therefore, proposed a much richer taxonomy of laboratory sciences, which he cleaved into three basic elements: marks

[1] Some of the key figures of this movement in the 1980s and early 1990s are Hacking (1983), Cartwright (1983, 1989, 1999), Franklin (1986), Galison (1987), Giere (1988) and Ackermann (1985). More recent important contributions have come from Mayo (1996) and Chang (2004).

(including observations, data and data-processing), matériel (including instruments and experimental procedures, and the substances or objects investigated) and ideas (including theories and models of instruments). He thereby rejects the idea that data can prove theories. Instead, Hacking argues that the elements that constitute laboratory practices are mutually adjusted. Laboratory practices eventually become stable because a proper fit has been established between these elements.

Siding with Hacking, I propose that different kinds of elements are mutually dependent and adjusted; they are stabilized in constructive activities by means of practical and theoretical reasoning, along with interventions that seek to explore and adjust their mutual interplay – which is an alternative to the traditional assumption that facts are unproblematic and serve to prove theories.

There is a third assumption related to the first and second traditional one: the idea that theories are science's most important results. However, if we accept the idea that the elements that constitute scientific practices, such as the ones distinguished in Hacking's taxonomy, are developed in a mutual interplay of interventions with the physical world (i.e. the natural world and technological devices and procedures) together with practical and theoretical reasoning, then these other elements must also be regarded as scientific results. In other words, it is not only theories that are the results of scientific research, but also data and data-processing, physical phenomena and their descriptions, instruments and their uses, experimental set-ups and technological procedures, scientific laws and models of data and phenomena, scientific methods, etc.

This suggestion is particularly significant for understanding scientific research in the context of application. Often, the purpose of these practices is to produce phenomena for practical uses, including the technological devices or procedures that bring them into being, and in tandem with practical and theoretical understanding of how these phenomena are produced (or prevented, controlled, improved, etc.). Moreover, in many cases the aim of these research practices is to create *artificial phenomena*. These phenomena are created by technological manipulation, for instance, in order to meet a certain technological function (cf. Boon and Knuuttila 2009). Engineering sciences, which is scientific research in the context of technological applications, is an example of a science in the context of application. Its purpose is scientific research that contributes to the development of technological devices, processes and materials. Usually, the proper (or improper) functioning of devices, processes and materials is understood in terms of phenomena that produce (or are detrimental to) their desired behaviour. By experimentation and scientific modelling, the engineering sciences strive to respectively understand and produce the specific behaviour of devices and processes and/or the properties of materials. In working towards this purpose, scientific practices develop three things in a mutual interplay: (1) experimental techniques and scientific instruments that enable the creation of and intervention with phenomena relevant to the functioning of technological applications; and (2) 'rule-like knowledge' and scientific models about (a) these phenomena and (b) how scientific instruments and experimental techniques produce the desired and undesirable phenomena.

To summarise, the third assumption, which holds that science is only interested in theories, is inadequate. Scientific practices, in particular those that work in the context of applications, produce different kinds of scientific results, which include data, physical phenomena, instruments, scientific methods and different kinds of scientific knowledge. In this dynamic, the fit between different kinds of scientific results is an important criterion for their acceptance. In this article I will focus on three aspects: the production and acceptance of physical phenomena as ontological entities; the role of instruments and experiments in their production; and the rule-like knowledge that is produced simultaneously.

Finally, a fourth (often implicit) assumption of traditional accounts is that theoretical knowledge somehow *represents* some kind of 'mind-independent' structure in the real world.[2] As an alternative position that entirely avoids accounts that involve the need for a representational relationship between epistemological results and the world, one might adopt Hacking's (1992) assumption that the stability of scientific results consists of a proper fit between different elements that constitute laboratory practices. This position circumvents the idea that our theories somehow represent a cognizable structure that exists in the world, independent of human ways of knowing. However, Hacking's notion of stability is not entirely satisfactory because it does not explain *why* a proper fit between these different elements leads to the *success* of science. In particular, part of the success of laboratory practices comprises an exchange of these elements among different practices. The fact that these elements seem capable of travelling independently of the laboratory context in which they were produced (also see Howlett and Morgan 2010) cannot be explained by 'the self-vindication of a laboratory practice'. As an alternative, I will propose (in Section 12.4), as a minimal metaphysical belief, the idea that the world is real (or *robust*) in the sense that it is external to us and stably sets limits to our interventions with it. This position is a kind of realism since it assumes that an independent real world sets limits to what we can *do* with it and to the regularities, causal relations, phenomena and objects that can possibly be determined. Yet, this kind of realism is minimal because it avoids the idea of a cognizable independent order or structure in the real world.

12.3 Part I: Conceptual Analysis of Robustness

12.3.1 Metaphysical, Regulative, Ontological, Methodological and Epistemological Robustness Notions

Wimsatt (1981, 2007) suggests that all the variants and uses of robustness share a common theme in distinguishing the *real* from the *illusory*; the *reliable* from the *unreliable*; the *objective* from the *subjective*; the *object* of focus from *artefacts* of

[2] Knuuttila and Boon (2011) present a critical analysis of how and why scientific models (and theoretical knowledge) give us knowledge. They argue that most philosophical accounts eventually draw on a representational relationship between scientific models and how the real world is.

perspective; and, in general, that which is regarded as ontologically and epistemologically *trustworthy* and *valuable* from that which is *unreliable*, *ungeneralizable*, *worthless* and *fleeting*. In this context, *robustness analysis* is of key importance to scientific methodology. Things or scientific results such as processes, laws and structures are reliable and valuable (or "robust") to their degree of invariance or stability under a robustness analysis, which involves the following procedures: analyzing a *variety* of *independent* derivation, identification or measurement processes; looking for and analyzing things which are *invariant* over or *identical* in the conclusions or results of these processes; determining the scope of the processes across which they are invariant and the *conditions* on which their invariance depends; and analyzing and explaining any relevant *failures of invariance*. Wimsatt calls these procedures *multiple-determination* or *robustness* (Wimsatt 2007, pp. 43–44).

Several other authors have also used 'robustness' and related notions to account for the epistemological or ontological character of scientific results, as well as for the way in which these results are accepted or justified. Pickering (1987, 1989) argues that scientific results are accepted, not because they correspond to something in the world, but because scientists bring so-called plastic resources in relations of mutual support, thus producing a "robust-fit" (cf. Hacking 1999). The resources are: the material procedure (including the experimental apparatus itself along with setting it up, running it and monitoring its operation); the theoretical model of that apparatus; and the theoretical model of the phenomena under investigation. As already mentioned, Hacking (1992) suggests that the results of mature laboratory science ('ideas, matériel and marks') achieve stability when the elements of laboratory science are brought into mutual consistency and support. Woodward (2001) seeks an account of the robustness of explanatory generalizations. He proposes that a generalization in biology is explanatory only if it is *invariant*, which means that it continues to hold under a relevant class of changes. Weisberg (2006) and Weisberg and Reisman (2008) argue that the *robustness* of theorems, such as the Lotka-Volterra principle that describes ecological processes, can be identified and confirmed by means of a robustness analysis (or stability analysis) of theorems.

Hence, robustness is used in the sense of reality, invariance, stability and reliability – other notions with a similar meaning are reproducibility, empirical adequacy and a notion that will be newly introduced in this context: 'same-conditions – same effects'. I will call them *robustness notions*. Interestingly, these robustness notions apply to different categories of things, such as physical processes and properties, scientific laws, theorems and models, methodological procedures and even the physical world or scientific practice as a whole. Indeed, despite their apparent synonymy, these robustness notions have distinct roles in the philosophical analysis of scientific practices. In order to account for these roles, I propose a conceptual distinction between *metaphysical*, *regulative*, *ontological*, *epistemological* and *methodological* robustness notions:

i. *Reality and stability* are properties of the physical world. We believe that the physical world is robust in the sense that it is external to us and stably sets limits to our interventions with it. In this context, reality and stability function as metaphysical robustness notions.
ii. *'Same conditions – same effects'* is a robustness notion that functions as a regulative principle of scientific practices. This principle says that under exactly the same physical conditions (in the natural world or in the laboratory or in technological devices) exactly the same physical effects will occur, which is an elementary assumption, the metaphysical or empirical truth of which cannot be proven. A regulative principle, therefore, is a fundamental presupposition or a 'condition of possibility' that facilitates scientific research: experimental sciences would not be possible without this presupposition. I will propose that 'same conditions – same effects' as a regulative principle is the philosophical basis of the other robustness notions and that it plays a crucial part in understanding the workings of these notions. This principle provides the condition for the possibility of metaphysical and/or logical principles that aim to justify inferences in scientific practices, such as induction, falsification and the *ceteris paribus* clause.
iii. *Reproducibility* denotes a property of measured data and observable or quantifiable physical occurrences that are produced by means of natural, experimental and/or technological conditions. Data and physical occurrences are considered as being reproducible if they are repeatable under the same technological and/or experimental conditions.
iv. *Stability* and *invariance* are ontological robustness notions because they are criteria for what can be accepted as *real* objects and phenomena. Importantly, scientists usually regard phenomena or objects as stable or invariant if they can intervene with them, for instance in experiments, or if they assume that they could intervene with them if they had better (or practically possible or ethically acceptable) procedures and technological means at their disposal (cf. Woodward 2003). Additionally, scientists accept that an object is real *because* it is invariant or stable when transferred to other circumstances, while they also accept that a phenomenon is real *because* it is invariant or stable in the sense that they can experimentally or technologically create, produce, control or even prevent its occurrence.
v. *Reliability* denotes a property of theoretical knowledge, such as phenomenological laws (or "rule-like" knowledge) and scientific models, in their epistemic use to create explanations and predictions about real-world situations. Reliability is an epistemological robustness notion because it is a criterion for accepting theoretical knowledge.
vi. *Empirical adequacy* denotes a property of fundamental theories such as Newton's or Maxwell's. It is an epistemological robustness notion because it applies to theoretical knowledge.
vii. *Repetition* and *multiple-determination* denote properties of scientific methods. They are methodological robustness notions that function as criteria for how scientific results such as reproducible data, stable phenomena and technological devices, reliable phenomenological laws and scientific models, empirically

adequate fundamental theories, etc., are produced and justified. These notions guide the development of scientific methodologies that warrant the production of 'robust' scientific results. Important aspects of multiple-determination are the above-mentioned aspects of a robustness analysis, e.g. variation, independence, invariance and failure of invariance (cf. Wimsatt 1981). The role of this methodological robustness notion is in the design and use of various technological instruments for producing measurements of the target system; the development of various technological devices and experimental procedures for the production of and intervention with phenomena; and the invention of methods for examining the proper and stable workings of these instruments and devices.

Hence, robustness notions function in different ways: Firstly, as a fundamental belief that we have about the (physical) world; secondly, as a regulative principle of scientific practices that justifies the functioning of other robustness notions and explains how these notions are related; thirdly, as a criterion for the actual existence of objects and phenomena; fourthly, as a criterion for the acceptance of epistemic results; and fifthly, as a criterion for methods that produce and justify the 'robustness' of measurements, phenomena and (theoretical) knowledge, i.e. methods that produce these scientific results and justify the attribution of epistemological and ontological properties. The proposed conceptual distinction between these robustness notions is summarized in Table 12.1.

Table 12.1 Conceptual distinctions of robustness notions

Category	Object	Robustness notion
Metaphysical	i. Reality→	i. Stable, deterministic, independent physical world
Regulative	ii. Scientific practice→	ii. 'Same conditions – same effects' as a presupposition or 'condition of possibility' for knowledge production
Ontological	iii. Measured data and physical occurrences→	iii. Reproducibility
	iv. Observable and theoretical objects, phenomena and causal relations→	iv. Stability and invariance
Epistemological	v. Phenomenological laws (or rule-like knowledge) and scientific models→	v. Reliability
	vi. Fundamental theories→	vi. Empirical adequacy
Methodological	vii. Scientific methods (that 'widen the span of phenomena, and the refinement of rule-like knowledge')	vii. Repetition and multiple-determination

12.3.2 Robustness as Truth-Maker or as an Alternative to Truth?

How do robustness notions relate to truth? Does 'multi-determination' taken as a methodological notion function as a 'truth-maker', or does reliability as an epistemological notion function as an alternative to truth? The two possibilities differ in the following way:

1. 'Robustness' functions as a *truth-maker*: Multi-determination functions as an alternative *methodological notion* that justifies the truth of scientific knowledge. It explains – or accounts for – *why* scientific knowledge is true. Scientific knowledge is true *if* it is the result of multiple-determination. This is how 'robustness' seems to function in Wimsatt (1981), Woodward (2001), Weisberg (2006) and Weisberg and Reisman (2008). 'Robustness' as a truth-maker is an alternative to how methodological notions such as 'induction', 'hypothetical-deduction' or 'inference to the best explanation', etc., are supposed to justify the truth of scientific knowledge.
2. 'Robustness' functions as an *alternative to 'truth'*: Reliability functions as an alternative *epistemological notion*. In other words, truth as the central property of scientific knowledge is substituted by reliability. In this account, multiple-determination may function as a methodological notion to justify the reliability of scientific knowledge but not its truth.

I will defend the second option and argue against the first. The questions to be answered then are: why and how robustness (in the sense of reliability) can function as an epistemological criterion, i.e. why and how it can function as an alternative to truth, and why robustness, in the sense of multi-determination, cannot function as a truth-maker.

Truth is an epistemological criterion. An epistemological criterion is the property that theoretical knowledge must have in order to be accepted or believed. This implies that the use of 'reliability' as an epistemological criterion must be similar to how 'truth' or other epistemological criteria, such as 'empirical adequacy', are used in statements such as: 'a theory is accepted *if* it is true' or 'a theory is accepted *if* it is empirically adequate'. Similarly, 'a theory is accepted *if* it is reliable'.

The crucial question is then how we know that a theory is true or empirically adequate or reliable. In other words, how do we justify that a theory has this epistemological property? In order to clarify this further, I will use Van Fraassen's (1980) well-known approach to the meaning and justification of the truth of scientific theories. First, I will outline his approach. Then, I will apply the resulting analytical schema to the analysis of 'reliability' as an alternative epistemological criterion.

Van Fraassen's point of departure is Tarski's semantic definition of truth, according to which truth is a property of a sentence, which tells us something about the relationship between the sentence and the real world. Van Fraassen's *definition* of the truth of a sentence or theory "T" is (slightly rephrased for my purpose): The truth of "T" means that what T says is literally the case – that is, "T" literally tells what the real world is like. Subsequently, a methodological criterion is needed that determines whether "T" literally says what the world is like. Van Fraassen's much-debated criterion is that the truth of statements or 'stories' can only be determined for the directly observable state of affairs and occurrences. In other words, the story told by "T" must be observable.

Following these ideas, I propose to explicate the use and meaning of epistemological criteria and how they relate to methodological criteria in five systematic steps:

1. *The epistemological criterion*. An epistemological criterion, E, (e.g. truth) accounts for the *acceptance* of theoretical knowledge "T". This criterion is used as follows: An expression "T" is accepted *if* "T" is E. In other words, an expression (e.g. a sentence or a scientific theory) called "T" and saying T is accepted if and only if the epistemological property (e.g. truth) has been attributed to the expression "T". For instance, a theory or law "T" (e.g. Newton's theory or the ideal gas law) is accepted *if* "T" is E (e.g. true).[3]
2. *A semantic conception of the epistemological criterion*. In this account, epistemological properties are regarded as semantic concepts. Semantic concepts deal with certain relations between expressions of a language and the object referred to by that expression (cf. Tarski 1944). This means that epistemological concepts are regarded as properties of expressions in a language, and not as properties of objects in the world to which these expressions refer. Accordingly, an epistemological property (e.g. truth) is a property of "T" (e.g. theoretical knowledge) that specifies a certain relationship between expression "T" and the real world.[4]
3. *A semantic definition of the epistemological criterion*. One of the characteristics of semantic concepts is that their meaning must be given by definition and not, for instance, by designation. Hence, a semantic definition of the epistemological property E must be given. The form of this definition is: An expression "T" is E means, or is defined as, that what T says relates such and such to the empirical world. For instance, that a theory or law "T" is true means that what T says is actually the case.
4. *An operational definition of the epistemological criterion*. One of the characteristics of concepts introduced by means of a definition rather than by means of

[3] Note that epistemological criterion E is a necessary property for scientific knowledge to be accepted, but may not be a sufficient criterion for acceptance, since other criteria, such as *relevance* or *explanatory power*, may play a role as well. Van Fraassen (1980, pp. 12–13) calls these additional criteria pragmatic values.

[4] In this manner, a distinction is made between properties of the world, e.g. material entities in the real world, and properties of expressions of a language, including theories. For example, red is regarded as a property of material or physical objects (e.g. the apple is red), whereas truth is regarded as a property of an expression (e.g. 'the apple is red' is true). Importantly, the way in which we learn their meaning is different. Usually, we learn the meaning of the properties of material objects by designation (e.g. by pointing at a red apple and saying 'Look! The apple is red.'), not by definition. The meaning of semantic concepts cannot be learned by designation (e.g. by pointing at something and saying 'Look! Newton's theory is true.'). Instead, the meaning of semantic concepts must be given by definition.

designation is that the use of that concept must also be defined.[5] This can be called an operational definition of the concept. The semantic definition of E (i.e. "T" is E means that what T says relates such and such to the empirical world) already includes the operational definition: "T" is E *if* what T says relates such and such to the empirical world. This latter version of the definition presents a criterion Q (e.g. is actually the case) for attributing the epistemological property E to a sentence "T". Namely, a sentence or scientific theory called "T" (and saying T about the empirical world) is E (e.g. true) if and only if what T says relates such and such to the empirical world. In short, the operational definition of the epistemological criterion reads: "T" is E (e.g. "T" is true) *if* T is Q (e.g. what T says about the empirical world is actually the case).

5. *The methodological criterion*. Hence, the problem of how to justify that the epistemic property for accepting theoretical knowledge applies (i.e. whether "T" is E) has been transferred to the problem of how to determine that T is Q (i.e. whether T relates such and such to the world). This is where methodology comes into play. Methodology involves a methodological criterion M (e.g. an observation), which is the quality a method must have in order to be accepted as a method by which it can be determined that T is Q. The use of this criterion is summarized as follows: "T is Q is justified *if* the question of whether T is indeed Q is determined by a methodology that meets methodological criterion M." For instance, the claim that 'what T says is actually the case' is justified *if* what T says can be directly observed in the real world. In short, observation counts as a methodological criterion: A method justifies the (approximate) truth of a sentence, a theory or a law *if* what the sentence, theory or law says is actually or literally the case. Whether, what the sentence, theory or laws says is literally the case, must be determined by observation.

In the case of truth as an epistemological property of theoretical knowledge, and direct observation as the methodological criterion for attributing this property to theoretical knowledge, this schema results in:

1. (1^T) *Epistemological criterion for the acceptance of theoretical knowledge*: "T" is accepted *if* "T" is true.
2. (2^T) *Semantic conception of the epistemological criterion*: Truth is an epistemological property of theoretical knowledge "T", which specifies a certain relationship between "T" and the real world, i.e. a relationship between what the theory says about the real world and how the world really or literally is.

[5] For instance, knowing how to use the term 'bachelor' (e.g. in saying, 'this man is a bachelor') requires an explication of how we determine whether this man is a bachelor. Similarly, in order to use a semantic concept such as truth in saying 'this theory or law is true', it needs to be explicated how we determine whether the theory is true. Importantly, a definition of a term (e.g. a definition of being a bachelor) not only states its meaning (e.g. a man is a bachelor means that a man is not married), it also presents a criterion for whether the term applies (e.g. a man is a bachelor *if* a man is not married).

3. (3^T) *Semantic definition of the epistemological criterion*: "T" is true means, or is defined as that what T says is actually the case.
4. (4^T) *Operational definition of the epistemological criterion*: "T" is true *if* what T says is actually the case.
5. (5^T) *Methodological criterion*: Direct observation is a methodological criterion for methods that determine whether what T says is actually the case. The use of this methodological criterion is summarized as follows: What T says is actually the case *if* 'what T says is actually the case' is determined by a methodology that is based on direct observation.

Clearly, if what our knowledge says about the world is observable in an unproblematic manner, we would not call it theoretical or scientific knowledge. Yet, the character of theoretical knowledge, "T", is that what T says is not observable in an unproblematic manner. According to Van Fraassen (1980), if T says something that is not observable in principle, we should refrain from attributing (approximate) truth to "T". In that case, this epistemological property does not apply and we need another property to account for, e.g. the acceptance or the success of "T". Van Fraassen proposed 'empirical adequacy' as an alternative notion, which is defined as: A theory "T" is empirically adequate *if* what it says about *observable* things in the world is true. Using this same line of reasoning, a methodological criterion is required to determine whether what the theory says about observable things is true. Van Fraassen (1980) and Suppe (1989) introduced the criterion of (partial) *isomorphism* between models that satisfy the axioms of the theory, on the one hand, and data models produced in experiments and data processing, on the other. In the case of empirical adequacy as an epistemological property of theoretical knowledge and (partial) isomorphism as the methodological criterion for attributing this property to theoretical knowledge, the former schema results in the following:

1. (1^{EA}) *Epistemological criterion for the acceptance of theoretical knowledge*: "T" is accepted *if* "T" is empirically adequate.
2. (2^{EA}) *Semantic conception of the epistemological criterion*: Empirical adequacy is an epistemological property of theoretical knowledge "T", which specifies a certain relationship between "T" and the real world (namely, a relationship between what the theory predicts about the observable world and what can be directly observed of the real world).
3. (3^{EA}) *Semantic definition of the epistemological criterion*: "T" is empirically adequate means, or is defined as that what T predicts about the observable world is actually the case.
4. (4^{EA}) *Operational definition of the epistemological criterion*: "T" is empirically adequate *if* what T predicts about the observable world is actually the case.
5. (5^{EA}) *Methodological criterion*: (Partial) isomorphism is a methodological criterion for methods that determine whether what T predicts about the observable world is actually the case. The use of this criterion is summarized as follows: What T predicts about the observable world is actually the case *if* 'what T predicts about the observable world is actually the case' is determined by a

methodology that is based on (partial) isomorphism, i.e. partial isomorphism between models that satisfy the axioms of the theory and data models of real-world systems (cf. Suppe 1989).

In summary, methodological criteria (i.e. direct observation of a state of affairs in the case of truth, and isomorphism between theoretical models and data models of a system in the case of empirical adequacy) are needed to justify the attribution of epistemological properties (i.e. truth and empirical adequacy) to theoretical knowledge.

This analysis reconstructs how Van Fraassen developed 'empirical adequacy' as the epistemological property that a theory must have in order to be accepted, which includes that empirical adequacy is proposed as an alternative to truth. Following this line of approach, aims to explore 'reliability' as an alternative epistemological criterion that accounts for the acceptance of theoretical knowledge in the scientific practices mentioned, instead of being a route to the truth of theoretical knowledge. Accordingly, the proposed schema results in the following:

1. (1^R) *Epistemological criterion for the acceptance of theoretical knowledge*: "T" is accepted *if* "T" is reliable.
2. (2^R) *Semantic conception of the epistemological criterion*: Reliability is an epistemological property of theoretical knowledge "T", which specifies a certain relationship between "T" and the real world, i.e. a relationship between what the theory predicts about the observable or measurable world and what can be observed or measured of the real world.
3. (3^R) *Semantic definition of the epistemological criterion*: "T" is reliable means, or is defined as that what T predicts about the empirical (observable or measurable) world is actually the case.
4. (4^R) *Operational definition of the epistemological criterion*: "T" is reliable *if* what T predicts about the empirical world is actually the case.
5. (5^R) *Methodological criterion*: Repetition and multiple-determination (cf. Wimsatt 1981) are methodological criteria for methods that determine whether 'what T predicts about the empirical world is actually the case'. The use of these criteria is summarized as follows: What T predicts is actually the case *if* 'what T predicts is actually the case' is determined by a methodology that is based on repetition and multiple-determination.

In summary, this analysis (in schema 1^R–5^R) proposes reliability as an alternative epistemological criterion for the *acceptance* of theoretical knowledge. Repetition and Wimsatt's (1981) notion of multiple-determination are proposed as criteria for methods that justify the attribution of reliability to theoretical knowledge. Multiple means of determination, according to Wimsatt, consists of using different sensory modalities to detect properties or entities; using different experimental procedures to verify empirical relationships or the existence of phenomena; using different assumptions, models or axiomatizations to derive theoretical results, etc. As a result, the schema draws a relationship between two different kinds of robustness notions: reliability (an epistemological criterion) is related to repetition and multiple-determination, which are methodological properties that a method must have in order to justify the reliability of theoretical knowledge.

At face value, the semantic definition of reliability as an epistemological criterion is the same as that of Van Fraassen's notion of empirical adequacy (compare 3^{EA} and 4^{EA} with 3^{R} and 4^{R}). I propose to distinguish between 'empirical adequacy' and 'reliability' as distinct epistemological criteria for different kinds of theoretical knowledge. In Van Fraassen's analysis, Newton's theory or Maxwell's theory, which have an axiomatic form, are used as examples. 'What the theory says' is understood as the scientific model that satisfies the axioms of the theory, such as the model of a harmonic oscillator and its theoretical data structures, e.g. curves in an x-t diagram. The theory is empirically adequate if these curves are (partially) isomorphic with the data structures produced by a real, but ideally behaving harmonic oscillator (cf., Suppe 1989). However, in many cases there is no abstract theory from which a model of the phenomenon can be deduced in a straightforward manner. In those cases, scientific models are theoretical interpretations of phenomena using different 'ingredients' (cf. Boon and Knuuttila 2009; Bailer-Jones 2009). In this case, 'what the theory says' is a theoretical interpretation of the phenomenon, which is accepted *if* it is reliable in explaining or predicting 'rule-like knowledge' produced by means of a variety of sufficiently independent experimental procedures and measurements (i.e. multiple-determination), for example.

Additionally, the difference between the two notions is related to different concepts of the epistemic aim of science, i.e. producing theories or producing epistemic results for specific epistemic purposes. Reliability as an epistemological property must account for the *use* of theoretical knowledge in performing epistemic tasks, such as in explaining or predicting specific phenomena in technologically produced circumstances. In other words, theoretical knowledge is reliable if it can perform the kind of epistemic tasks for which the knowledge is produced.

Provisionally, I propose to use empirical adequacy as an epistemological criterion for theories that have an axiomatic form – and which are usually called 'fundamental theories' –where reliability applies to theoretical knowledge that has as its primary aim the reliable (mathematical or verbal) description, explanation or prediction of phenomena (see also Table 12.1).

Based on this analysis, I will conclude that the acceptance of theoretical knowledge does not necessarily run via truth. I will also adopt Van Fraassen's critical point that truth only applies to descriptions of a state of affairs that can be directly observed in an unproblematic manner, where truth is inappropriate as an epistemological property of theoretical knowledge.[6]

[6] In accepting Van Fraassen's claim, I deliberately ignore the well-known critique with regard to his notion of observability. The important point of Van Fraassen's suggestion is, in my view, that we have a more or less intuitively clear understanding of the meaning of truth in every day situations. In those situations, we know how to use this notion and how it functions in distinguishing

However, it still needs to be explained *why* a methodological criterion justifies the attribution of an epistemological criterion, i.e. why theoretical knowledge produced by means of repetition and multiple-determination is reliable or, in line with the proposed conceptual structure, why do scientific methods with this methodological criterion as a property justify that theoretical knowledge is reliable? This question will be addressed in Section 12.4 of my argument.

12.4 Part II. How Robustness Notions Work Together as Criteria for Producing Scientific Results

12.4.1 'Same Conditions – Same Effects' as a Regulative Principle

The physical world is real or robust in the sense that an independent world stably sets limits to what we can do with it and to the regularities, causal relations, phenomena and objects that can possibly be determined. This metaphysical idea functions in scientific practices by way of the assumption that with exactly the same physical conditions exactly the same physical effects will occur. This belief involves a metaphysical principle about 'how the physical world is', which reads: There is one stable, deterministic physical world in which the same physical conditions will always produce the same physical effects.[7] The philosophical problem of metaphysical principles is that there is no method to prove them, e.g. to find out whether the same conditions will always produce the same effects.[8] At the same time, the belief that the world is structured, regular or stable inescapably 'regulates' our interactions with and our thinking about the world. It is a belief without which thinking about the world and producing knowledge that guides our thinking and acting would be impossible. Therefore, I propose to regard 'same conditions – same

---

between claims that are true and those that are not. The use of this notion with regard to theoretical knowledge, on the other hand, is not intuitively clear.

[7] Stochastic behaviour in quantum physical experiments does not violate the idea of 'same conditions – same effects' given that under the same conditions the same stochastic behaviour will occur. Hence, physicists still work with the presupposition that the same experimental set-up in quantum physics will produce the same patterns.

[8] This problem resembles David Hume's problem of causal relations: How can we know that causes and effects will be related in the future as they were in the past if we cannot find out empirically which *power, force, energy* or *necessary connexion* keeps them together? (Hume (1777 (1975)). *On the Idea of Necessary Connexion*, Part I in: *Enquiry concerning Human Understanding.*) Accordingly, Hume framed the problem of inferring a stable relationship between cause and effect as a fundamental problem of empiricism: we cannot *observe* the connection in an unproblematic manner – as a consequence, inductive inference to a stable relationship between cause and effect cannot be empirically verified. To this fundamental problem of empiricism, Popper (1963) added that inductive inference cannot be logically justified either. In order to avoid such metaphysical problems, Popper framed it as the problem of induction, i.e. as a problem of the logic of science. The underlying philosophical problem is that the metaphysical belief that the world is structured, regular or robust cannot be proven.

effects' as a *regulative principle* that 'guides and enables' the production and justification of knowledge about the world, by means of which we think about and act in it.[9] A regulative principle is one that scientists must adopt in order to enable scientific and practical reasoning about the world, while at the same time they must acknowledge that it is not possible to find out whether this principle is an empirical or metaphysical truth.

In my view, 'same conditions – same effects' as a regulative principle that 'guides and enables' scientific inferences is more appropriate as an account of how and why 'robust' knowledge about the real world is possible than logical principles, e.g. the principle of induction or falsification or the *ceteris paribus* clause, or metaphysical principles, e.g. the principle that there must be a conceivable independent order or structure in the world (see also note 8). It is more appropriate in the sense that it accounts for the refined way in which scientific practices actually produce, justify and use knowledge.

Importantly, in scientific practices, we do not know what exactly belongs to the conditions nor do we have complete knowledge of what belongs to the effects. Scientists usually have to find out what the (causally relevant) conditions are and what the relevant effects are. Accordingly, this principle guides what scientists should look for (to wit, phenomena and the conditions that are causally relevant to their occurrence or existence or deterioration) and it justifies inference to general rules of the form: 'If A then B provided C, unless other causally relevant conditions K (known) and/or X (unknown)', rather than, 'If A then B'. Hence, the general rules that are justified by 'same conditions – same effects' are conditional. They enable and justify predictions in new situations, while simultaneously stating that new situations may involve other (known or unknown) causally relevant conditions that affect the phenomenon.[10]

---

[9] Understanding metaphysical presuppositions as regulative ideas was Kant's solution to the problems of empiricism raised by Hume. I do not claim that 'same conditions – same effects' is the only kind of basic belief that enables and guides scientific research. I largely agree with Chang (2009), who, with a similar Kantian approach, aimed to explain the functioning of these kinds of principles. He proposed calling basic beliefs about the world 'ontological principles', which – similar to what I claim about the function of 'same conditions – same effects' as a regulative principle – enable epistemic activities such as observation, experimentation, counting, logical reasoning, etc.

[10] 'Same conditions – same effects' differs from the *ceteris paribus* clause in the sense that the latter does not count 'all other conditions' as part of the rule-like knowledge, whereas the former counts any addition to knowledge of them as an extension of the rule-like knowledge. In scientific practice, this difference is crucial because explicit knowledge of these conditions (C_{device} and K) enables us to predict under which circumstances the phenomenon described as $A \rightarrow B$ can or cannot be expected. Ceteris paribus laws only apply to what Cartwright calls a nomological machine: the law applies only when 'all other conditions being equal,' which would only allow for a very limited use.

12.4.2 Reproducibility and Stability as an Ontological Criterion for the Acceptance of Phenomena

'Same conditions – same effects' as a regulative principle points to a different idea about the nature of phenomena than the commonly accepted ideas, such as those articulated by Hacking (1983), Bogen and Woodward (1988) and Bailer-Jones (2009).[11] Contrary to what philosophers often suggest, phenomena are usually not the point of departure of scientific research. Identification and reproducible technological (or experimental) production of physical phenomena is a central activity of scientific practices, in particular of those practices that are conducting research in the context of application. What is essential to my account of 'same conditions – same effects' is that phenomena themselves must be recognized as technological achievements, as well as ontological and epistemological achievements.

In order to appreciate these claims, the nature of phenomena needs to be explained a bit further. Common language suggests that phenomena must be regarded as independent, 'freely floating' physical entities. Sentences such as 'we observe a phenomenon' or 'we isolate a phenomenon by means of a technological device' suggest that phenomena are very much like the grains of sand on a beach or heavenly bodies in an empty space. However, phenomena do not exist as isolated objects (see also Trizio 2008). They exist, emerge or disappear under specific physical conditions. In other words, phenomena are usually determined by and are dependent on physical conditions and, in principle, they can interact with or be affected by any other physical condition, thereby producing a different phenomenon. For this reason, and as explained above, an infinite number of physical phenomena can, in principle, be identified (see also McAllister 1997, 2011).

As a consequence, 'simple' phenomena must be regarded as ontological entities that are physically 'carved-out' by us. A 'simple' phenomenon is constrained by how the physical world is, but shaped into something by experimental interventions and/or technological devices and (formally) described as A→B; for instance, the phenomenon described by, "If gas is heated (A), it expands (B)."[12,13] Usually, identifying and describing them also involves pragmatic considerations. To be considered as an ontological entity requires that a phenomenon is regarded as

[11] Hacking (1983, p. 221) is canonical: 'A phenomenon is noteworthy. A phenomenon is discernible. A phenomenon is commonly an event of process of a certain type that occurs regularly under definite circumstances'.

[12] The understanding of phenomena I propose can loosely be explained by an analogy with Aristotle's notion of four causes of an object: the physical world is the material cause of a phenomenon, whereas our technological devices and experimental set-up are their formal cause. Additionally, the scientist is the efficient cause for describing the phenomenon as A→B, while the scientific or practical purpose for which the phenomena described as A→B is 'carved out' is its final cause.

[13] Clearly, some phenomena are observable in principle, e.g. the orbits of planets, the tides, an apple falling. However, as Kant has already argued, we are already actively involved even in 'simple' observations of phenomena, i.e. we actively 'carve them out' even in 'simple' observations. Massimi's article (2008) on this matter is insightful.

(qualitatively) relevant and/or (quantitatively) significant for one purpose or another, such as for understanding the behaviour of a specific physical system or for being used in technological applications. Only then does a phenomenon acquires ontological status, and as such becomes an ontological achievement. Furthermore, the phenomenon described as A→B is an epistemological achievement. In order to express this entangled ontological and epistemological understanding of physical phenomena, I propose to use the expression 'the phenomenon described by A→B', rather than, 'the phenomenon P' or 'the phenomenon A→B'.

Experimental interventions with technological devices will also produce knowledge of conditions that are causally relevant to the reproducible production of a phenomenon described by A→B, which is presented in 'rule-like' knowledge in the form: $A + C_{device} \rightarrow B$, unless (K and/or X). For instance, in experimental interventions with a particular device, e.g. heating a gas in a gas-tight cylinder with a freely moving piston, it has been found that 'if A then B', e.g. if gas is heated, then it expands. Additionally, it has been examined how the working of the device contributes to this phenomenon, resulting in a description of the causally relevant conditions of the device, C_{device}, e.g. the device contains the gas and allows for its free movement. In this manner, rule-like knowledge has been produced in the form "if A then B, provided C_{device}, unless other known (K) and/or yet unknown (X) causally relevant conditions."

The question that still has to be answered is how the acceptance of a phenomenon described as A→B works: how does a phenomenon acquire ontological status?

In scientific practices, *reproducibility* applies to measured data and observed physical occurrences, which are either naturally or technologically produced. Subsequently, reproducible physical occurrences may acquire ontological status and thus be referred to as phenomenon described as A→B. It will usually acquire ontological status only if a physical occurrence that reproducibly appears in a specific set-up also occurs at other (technologically produced) circumstances. If not, we merely have an occurrence and/or a measured data-set that is reproducibly produced by that specific device. In other words, in order to acquire ontological status, the physical occurrence must be *stable or invariant* in the sense that it occurs when the same conditions occur at other (technologically produced) circumstances, e.g. another kind of technological device or experimental set-up.

In analyzing how phenomena described as A→B are produced and justified, I propose that this also involves ontological and methodological robustness notions playing a role, similar to how these notions play a role in the acceptance of epistemic results. I will take *reproducibility and stability or invariance* as a combined ontological robustness notion, although a more refined analysis should separate them. Accordingly, the formerly proposed analytical schema results in the following:

1. (1^O) *Ontological criterion for the acceptance of a phenomenon*: A phenomenon described as A→B is accepted as real *if* it occurs reproducibly and is stable or invariant.
2. (2^O) *Semantic conception of the ontological criterion*: Reproducibility and stability or invariance is a property of a phenomenon described as A→B, which

specifies a certain relationship between the description, A→B, and an occurrence in the real world, i.e. a relationship between what this description A→B says about the real world and an occurrence that happens in the real world.

3. (3^{O}) *Semantic definition of being reproducible and stable*: a phenomenon described as A→B is reproducible and stable means (or is defined as) that the same conditions, A+C_{device}, will produce the same effects, B, unless (K and/or X).
4. (4^{O}) *Operational definition of* being reproducible and stable: a phenomenon described by A→B is reproducible and stable *if* the same conditions, A+C_{device}, will produce the same effects, B, unless (K and/or X).
5. (5^{O}) *Methodological criterion*: Repetition and multiple-determination are methodological criteria for methods that justify the reproducibility and stability of phenomena described by A→B, as well as the reliability of rule-like knowledge of the form: 'same conditions A+C_{device}, will produce the same effect B, unless (K and/or X)'. Hence, a phenomenon described by A→B is reproducible, stable and invariant (for use in practical applications) *if* the rule '$A + C_{device} \rightarrow B$, unless (K and/or X)' has been produced and justified by multiple-determination. Conditions $K_1 \ldots K_n$ that are causally relevant for the phenomenon described as A→B under other relevant circumstances must be determined by repetition and multiple-determination.

To summarize, the proposed schema relates reproducibility, stability and invariance as ontological criteria for the acceptance of a phenomenon described as A→B to multiple-determination as a methodological criterion for justifying this attribution of ontological status. This schema also shows that referring to a phenomenon as an ontological entity is intertwined with the experimentally produced rule-like knowledge about it.

Finally, it needs to be explained *why* repetition and multiple-determination justify the acceptance of a phenomenon described as A→B and of the rule-like knowledge about it, or, in line with the proposed conceptual structure, why do scientific methods that have these methodological criteria as a property justify that a phenomena described as A→B is *reproducible* and *stable*, and that rule-like knowledge in the form '$A + C_{device} \rightarrow B$, unless (K and/or X)' is *reliable*?

12.4.3 Repetition and Multiple-Determination as Methodological Criteria for Justifying the Acceptance of Scientific Results

A central idea of my analysis is that epistemological and/or ontological properties can only be attributed to a scientific result by methodological criteria that justify this inference. Thus, in traditional philosophical accounts, induction or hypothetical-deduction or 'severe tests' are supposed to function as methodological criteria that justify inference to the truth or empirical adequacy of a scientific theory. One of the tasks of the philosophy of science is to give an account of *why* a methodological criterion justifies this inference.

Wimsatt (1981) proposed that robustness *is* multiple-determination. However, he is not fully clear about what is achieved by robustness as a methodological criterion (in my terminology). One can argue that multiple-determination functions as an alternative methodological criterion that justifies inference to the (approximate) *truth* of epistemic results and the *reality* of ontological results. Alternatively, Wimsatt may have meant that multiple-determination functions as a methodological criterion for producing 'robust' (rather than true or real) results, which is in line with my own proposal in this article. In both cases, an account is needed of why multiple-determination as a methodological criterion justifies that a scientific result is (approximately) true, real or 'robust'.

I will argue that multiple-determination cannot justify the attribution of (approximate) truth to theoretical knowledge nor independent reality to phenomena described as A→B. Instead, I propose that in scientific practice the character of accepted phenomena and scientific knowledge is much more moderate and refined. In my proposal, these kinds of scientific results are accepted *because* ontological or epistemological robustness notions apply to them. Three aspects of 'same conditions – same effects' are important for understanding what exactly has been achieved (if not 'truth') by attributing these properties.

Firstly, 'same conditions – same effects' as a regulative idea may incorrectly suggest that *repetition* is sufficient as a methodological criterion for justifying the acceptance of phenomena. Repetition as a methodological criterion would work as follows: Data and physical occurrences are reproducible and stable *if* they are the same in every repetition. More specifically, according to the proposed conceptual schema, the fifth statement would then read:

5. ($5^{\text{O-false}}$) Repetition is a methodological criterion for methods that justify the reproducibility and stability of phenomena described as A→B, together with reliable rule-like knowledge in the form 'same conditions $A+C_{device}$ will produce the same effect B, unless (K and/or X).'

Indeed, in scientific practices the reproducibility of data and physical occurrences produced in measurements and experimental procedures is (partly) justified by repetition.[14] Yet, repetition is insufficient as a methodological criterion for justifying the stability of phenomena described as A→B because repetition does not present us with knowledge of causally relevant conditions, C_{device} and K, that must be created or prevented to produce the phenomenon described as A→B.

Multiple-determination is a methodological criterion that goes beyond mere repetition. In experiments, the variable conditions (e.g. temperature, pressure,

[14] In scientific practices, repetition is often too limited as a methodological criterion for reproducibility because repetition (or replication) often does not produce the same results. This is because not all relevant causal conditions are known. If repetition shows anomalous behaviour, a possible but not entirely essential explanation is that the previously measured data or phenomena are not reproducible and were, therefore, artefacts. Usually, scientists will search for 'hidden' causally relevant conditions.

speed, size, chemical concentrations, fluidic movements and electro-magnetic field strength) of the natural environment, the technological means, and/or the experimental procedure are varied in order to see what happens to the phenomena described as A→B produced by these systems. Using this approach, scientists find out whether phenomena described as A→B are causally influenced by these conditions and how sensitive they are to them (see also Woodward 2003). This methodological approach is also called a *sensitivity analysis*.

Multiple-determination also involves other instruments or procedures being employed either to examine the proper functioning of the equipment or to expand on the conditions under which measurements or experiments are performed. This can be done, for instance, by enhancing the sensitivity (e.g. the sensitivity of measuring a variable parameter by using other types of instruments) or in range (e.g. the range of values of a variable condition is enlarged in order to see the effects at the limits) or in complexity (e.g. other kinds of phenomena are simultaneously produced in order to see whether they affect the phenomenon described as A→B). Wimsatt (1981, 2007) and Franklin (1986, 2009) have listed a wide variety of strategies that illustrate ways in which scientific practices employ 'multiple-determination' to examine the robustness of scientific results (including technological devices and experimental procedures).

Secondly, the point of repetition as a methodological criterion is to produce the *same* results (thus, strictly, the same data and the same physical occurrences), whereas the point of multiple means of determination is that it usually does not produce the same results, at least not at the level of our observations or measurements. Yet, in his examples of multiple-determination Wimsatt suggests that the point of it is producing the *same* results: '. . .to detect the *same* property or entity', '. . .to verify the *same* empirical relationships or generate the *same* phenomenon', etc. (Wimsatt 2007, p. 45, my italics). This way of phrasing how multiple-determination works suggests, again, that phenomena are like grains of sand on the beach. They are clearly identifiable objects in whatever circumstances: they remain as exactly the same identifiable entities whether on the beach, at the bottom of the sea, in the belly of a fish or in my shoes. As I have explained, this is often not the case for phenomenon described as A→B under new circumstances. The point of 'same conditions – same effects' as a regulative principle is that scientists must seek to discover conditions that occur under other circumstances, and whether these conditions are causally relevant to the phenomenon described as A→B, and also how these conditions account for results that deviate from the phenomenon described as A→B. As a consequence, multiple-determination is a methodological criterion for determining conditions that are causally relevant to phenomena described as A→B and for determining how sensitive phenomena are to the causally relevant conditions.

Thirdly, a related aspect of 'same conditions – same effects' is that a phenomenon described as A→B is stable – and the rule-like knowledge about it is reliable – to 'some extent' or 'conditionally'. The extent to which these results are stable or reliable, respectively, is given by the extent to which they have been put to experimental tests. This conditional character of scientific results has been made operational in the following manner. The regulative principle 'same conditions – same effects' guides and enables scientists to carve out phenomena described as A→B, accompanied by the production of rule-like knowledge in the form: '$A + C_{device} \rightarrow B$, unless (K and/or X).' Repetition and multiple-determination characterize the methodological approach by which K has been found, and by which scientists continue to search for X. Accordingly, repetition and multiple-determination are methodological criteria that account for the fact that the reliability of scientific knowledge lies in the *span* of simple phenomena described as A→B, and the *refinement* of rule-like knowledge about these phenomena, which can only be acquired by experimental interventions with the natural world and with technological instruments, devices and procedures.

Based on this analysis, the crux of the role that robustness notions play in scientific practices can be summarized. A central aim of traditional philosophical accounts was to justify methodologies by which we could possibly infer the truth of theoretical knowledge and/or the real and independent existence of theoretical objects. Based on the analyses by means of the proposed conceptual schema, I suggest that attributing these highly desired epistemological and ontological properties to scientific results *transcends* the methodological and regulative criteria that are the leading factor in producing and accepting them. Robustness as a truth-maker would work as follows: theories are true *because* we found that they are robust, i.e. reliable. Similarly, phenomena described as A→B exist independently *because* we found that they are robust, i.e. stable or invariant. The point of my argument against 'robustness' as a truth-maker is that attributing epistemological and ontological properties that transcend what has been attained by means of the methodological and regulative criteria is philosophically problematic.

As an alternative, I propose that robustness notions work together in a manner that avoids this kind of transcendence. Regulative, methodological and epistemological or ontological criteria are used in a mutual interplay, thereby guiding and enabling the production and acceptance of scientific results. 'Same conditions – same effects' is a regulative principle that justifies repetition and multiple-determination as methodological criteria for producing results that are defined as epistemologically or ontologically robust. Accordingly, it is justified to accept that: 'the phenomena are reproducible and stable *if* they have been determined by repetition and multiple-determination', while it is unjustified to conclude that, 'the phenomena are reproducible and stable, and therefore they exist independently.' Similarly, the following inference is justified: 'rule-like knowledge is reliable *if* it has been determined by repetition and multiple-determination', whereas, 'rule-like knowledge is reliable (or robust, cf. Weisberg and Reisman 2008), and therefore it is true,' is unjustified. In brief, transcendence to the highly desired epistemological and ontological properties by means of methodology cannot be justified. By using methodological robustness notions, robust (stable and reliable) scientific results are produced – nothing more and nothing less. As a consequence, the idea that 'robustness' is a 'truth-maker' must be rejected.

This restriction of scientific inferences is important to gain a better understanding of what science can do and what not. Science is much more limited than philosophers of science tend to believe. It must be kept in mind that the stability of a phenomenon described as A→B, and the reliability of the rule-like knowledge that accompanies it, is only justified to the extent that it has been put to the test. This account has been made operational by stating that the rule-like knowledge is conditional in the form: '$A + C_{device} \rightarrow B$, unless (K and/or X).' Hence, the proposed account of robustness notions is more appropriate for scientific practices than common traditional accounts of the justification of scientific results. Unjustified transcendence is avoided because the regulative principle and methodological criteria for producing and accepting a scientific result *define* the meaning of the epistemological or ontological property that is attributed to scientific results, which implies that scientific results are accepted *because* they have this epistemological or ontological property.[15]

12.5 Conclusions

The general scope of my interest in robustness notions is to understand scientific practices that work in the context of practical and technological applications. How can we explain their successes and limits? What can these practices do and what can they not do? Do we have to explain the applicability of scientific results using their truth? In this article, I have developed an account of robustness notions that may provide us with a more appropriate understanding of these practices. My argument aims to make plausible that explaining the success of scientific practices does not necessarily happen via the truth of scientific theories and/or the independent existence of theoretical entities, since 'robustness', as it is interpreted here, can sufficiently explain what science can do, while it also explains why science is limited. Here, I will summarize the structure of my argument.

In order to create a philosophical space within which the issues mentioned can be analysed, I have proposed four philosophical presuppositions as alternatives to some of the dominant traditional ones that tend to make important aspects of these scientific practices invisible or turn them into 'non-issues'. These alternative presuppositions were used as the philosophical foundation for understanding the different kinds of roles of robustness notions in scientific practices that produce, justify and use scientific results. They can be summarized as follows: (a) The *epistemic* aim of science is to produce epistemological results that allow for scientific reasoning about the world; (b) Scientific practices employ a methodology in which different kinds of elements are mutually adjusted and stabilized; (c) These different kinds of elements must each be recognized as different kinds of scientific results; (d) 'Same conditions – same effects' is an essential presupposition without which experimental practices cannot function or, in other words, it is a regulative principle that makes these practices possible since it guides the production and justification of empirical results.

[15] Which resonates with one of the central ideas of logical positivism that the meaning of a synthetic statement is the method of its empirical verification.

My focus is on scientific research in the context of practical and technological applications. In that context, epistemic results, such as scientific theories, models and concepts, but also rule-like knowledge about phenomena described as A→B, are *accepted* not necessarily because they are true, but because they enable and guide our thinking about the world and/or about intervening with it, in a relevant and reliable manner. As a consequence, a philosophical account is needed of how scientific results that meet this epistemic function are justified.

I have proposed a conceptual schema for analyzing the acceptance of epistemological results. The development of this schema was motivated by Van Fraassen's (1980) analysis of true scientific knowledge, which is founded on the following ideas: (i) Truth is an epistemological property of knowledge, not of the world; (ii) Knowledge is accepted *because* it has this epistemological property; (iii) The attribution of an epistemological property must be justified by a methodological criterion. I adopt Van Fraassen's idea that truth is inappropriate as an epistemological property because what the theoretical knowledge describes cannot be observed in a straightforward manner. Epistemological properties other than truth, e.g. empirical adequacy or robustness, may justify the acceptance of theoretical knowledge.

Based on an analysis of how several authors in the philosophy of science have used robustness in accounting for the success of science, I have proposed a conceptual distinction between *metaphysical, regulative, methodological, ontological and epistemological* robustness notions. These notions *function* as properties and criteria for different kinds of things. Reality and stability function as a metaphysical robustness notion about how the world is. Reproducibility and stability function as an ontological criterion for the acceptance of data and phenomena. Reliability functions as an epistemological criterion for the acceptance of scientific knowledge, while repetition and multiple-determination function as methodological criteria for the production and justification of epistemological and ontological results. The notion 'same conditions-same effects' is introduced as a regulative robustness notion. Next, the proposed conceptual schema is utilized to explain how these different robustness notions are related in the production and acceptance of scientific results.

Following Hacking's view that the stability of experimental sciences results from the mutual adjustment of different kinds of elements, thereby producing a self-vindicating structure, I have suggested that in order to explain why scientific results can travel to other scientific fields or technological applications, some kind of realism is needed. As a minimal metaphysical belief, I proposed that the real world is stable and independent in the sense that it puts real constraints on what we can *do* with it – what we can *think* about the real world, on the other hand, is constrained but *not* determined by it.[16] This metaphysical belief claims that the same physical conditions will always produce the same physical effects, but does not claim that there is an independent *cognizable* order or structure in the world.

[16] This realism is close to Hacking's realism, which emphasizes the materiality of the world.

What is crucial to my argument is that the regulative principle 'same conditions – same effects' explains and justifies *why* methodological criteria (multiple-determination) justify the acceptance of epistemological and ontological results. This argument, which expands on Wimsatt's account of the methodological role of multiple-determination, explains the appropriateness of the methodological criteria of repetition and multiple-determination for producing and justifying reliable rule-like knowledge that is also conditional. Multiple-determination also accounts for the fact that the reliability and relevance of scientific results lies in the *span* of simple phenomena described as A→B, and the *refinement* of rule-like knowledge about these phenomena, which can only be acquired by varying (mutually independent) experimental interventions with the natural world or with technological instruments, devices and procedures. This account also implies that the regulative principle 'same conditions – same effects' is more appropriate as scientific inference than logical principles such as induction, falsification or the *ceteris paribus* clause.

Finally, this account leads to the conclusion that robustness is not a truth-maker, i.e. multiple-determination cannot function as a methodological criterion for justifying that theoretical knowledge is true. The crucial point of the latter argument is that epistemological and ontological properties of scientific results cannot transcend the methodological criteria that led to their production and acceptance.

Acknowledgements I would like to thank Léna Soler and the PractiScienS group for their agenda-setting endeavours on this topic and for their suggestions for improving this text. I would also like to thank Henk Procee for his numerous suggestions on the topic and on the content of this chapter. This research is supported by a Vidi grant from the Dutch National Science Foundation (NWO).

References

Ackermann, R.J. 1985. *Data, Instruments and Theory: A Dialectical Approach to Understanding Science*. Princeton, NJ: Princeton University Press.

Ackermann, R. 1989. "The New Experimentalism." *The British Journal for the Philosophy of Science* 40(2):185–90.

Bailer-Jones, D.M. 2009. *Scientific Models in Philosophy of Science*. Pittsburgh, PA: Pittsburgh University Press.

Bogen, J., and J. Woodward, 1988. "Saving the Phenomena." *The Philosophical Review* 97(2):303–52.

Boon, M. 2006. "How Science is Applied in Technology." *International Studies in the Philosophy of Science* 20(1):27–47.

Boon, M. 2009. "Understanding in the Engineering Sciences: Interpretative Structures." In *Scientific Understanding: Philosophical Perspectives*, edited by Henk W. de Regt, Sabina Leonelli, and Kai Eigner, 249–70. Pittsburgh, PA: Pittsburgh University Press.

Boon, M. 2011. "In Defence of Engineering Sciences. On the Epistemological Relations Between Science and Technology." *Techné: Research in Science and Technology* 15(1):49–71.

Boon, M., and T.T. Knuuttila 2009. "Models as Epistemic Tools in Engineering Sciences: A Pragmatic Approach." In *Handbook of the Philosophy of Science. Volume 9: Philosophy of Technology and Engineering*, edited by Anthonie Meijers, 687–720. Amsterdam: Elsevier.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press, Oxford University Press.

Cartwright, N. 1989. *Nature's Capacities and their Measurement*. Oxford: Clarendon Press, Oxford University Press.

Cartwright, N. 1999. *The Dappled World. A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Chang, H. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Chang, H. 2009. "Ontological Principles and the Intelligibility of Epistemic Activities." In *Scientific Understanding: Philosophical Perspectives*, edited by Henk W. de Regt, Sabina Leonelli, and Kai Eigner, 64–82. Pittsburgh, PA: Pittsburgh University Press.

Franklin, A. 1986. *The Neglect of Experiment*. Cambridge, NY: Cambridge University Press.

Franklin, A. 2009. "Experiment in Physics." *The Stanford Encyclopaedia of Philosophy* (Spring 2009 Edition), Edward N. Zalta, (ed.), http://plato.stanford.edu/archives/spr2009/entries/physics-experiment/.

Galison, P. 1987. *How Experiments End*. Chicago, London: University of Chicago Press.

Giere, R.N. 1988. *Explaining Science*. Chicago and London: The University of Chicago Press.

Hacking, I. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hacking, I. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, edited by A. Pickering, 29–64. Chicago: University of Chicago Press.

Hacking, I. 1999. *The Social Construction of What*? Cambridge, MA: Harvard University Press.

Howlett, P., and M.S. Morgan, eds. 2010. *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*. London: Cambridge University Press.

Hume, D. (1777 (1975)). *Enquiries concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press.

Knuuttila T.T., and M. Boon. 2011. "How do Models Give Us Knowledge – The Case of Carnot's Ideal Heat Engine." *European Journal Philosophy of Science* 1(3):309–334.

Massimi, M. 2008. "Why There Are No Ready-Made Phenomena: What Philosophers of Science Should Learn From Kant." *Royal Institute of Philosophy Supplement* 63:1–35.

Mayo, D.G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

McAllister, J.W. 1997. "Phenomena and Patterns in Data Sets." *Erkenntnis* 47:217–28.

McAllister, J.W. 2011. "What do Patterns in Empirical Data Tell Us About the Structure of the World?" *Synthese* 182(1):73–87.

Morrison, M., and M.S. Morgan. 1999. "Introduction." In *Models as Mediators – Perspectives on Natural and Social Science*, edited by M.S. Morgan and M. Morrison, 1–9. Cambridge: Cambridge University Press.

Pickering, A. 1987. "Constructing Quarks. 'Against Correspondence: A Constructivist View of Experiment and the Real' ". In *PSA 1986*, vol. 2, edited by A. Fine and P. Machamer, 196–206. Pittsburgh: Philosophy of Science Association.

Pickering, A. 1989. "Living in the Material World: On Realism and Experimental Practice." In *The Uses of Experiment*, edited by D. Gooding, T. Pinch, and S. Schaffer, 275–97. Cambridge: Cambridge University Press.

Popper, K.R. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.

Rouse, J. 2009. *Articulating the World: Toward a new Scientific Image*. Contribution at San Francisco State Workshop: *The Role of Experiment in Modeling*. March 2009. http://wesfiles.wesleyan.edu/home/jrouse/Articulating%20the%20World.pdf

Rouse, J. 2011. "Articulating the World: Experimental Systems and Conceptual Understanding." *International Studies in the Philosophy of Science* 25(3):243–255.

Suppe, F. 1989. *The Semantic Conception of Theories and Scientific Realism*. Urbana and Chicago: University of Illinois Press.

Tarski, A. 1944. "The Semantic Conception of Truth: And the Foundation of Semantics." *Philosophy and Phenomenological Research* 4(3):341–76.

Trizio, E. 2008. "How Many Sciences for One World? Contingency and the Success of Science." *Studies in the History and Philosophy of Science* 39:253–8.

Van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Weisberg, M. 2006. "Robustness Analysis." *Philosophy of Science* 73(5):730–42.

Weisberg, M., and K. Reisman. 2008. "The Robust Volterra Principle." *Philosophy of Science* 75(1):106–31.

Wimsatt, W.C. 1981. "Robustness, Reliability, and Overdetermination." In *Scientific Inquiry in the Social Sciences*, edited by M. Brewer and B. Collins, 123–62. San Francisco, CA: Jossey-Bass.

Wimsatt, W.C. 2007. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA and London: Harvard University Press.

Woodward, J. 2001. "Law and Explanation in Biology: Invariance is the Kind of Stability that Matters." *Philosophy of Science* 68(1):1–20.

Woodward, J. 2003. *Making Things Happen, A Theory of Causal Explanation*. Oxford: Oxford University Press.

Chapter 13
The Robustness of Science and the Dance of Agency

Andrew Pickering

'Robustness' can mean many things. Wimsatt's (1981) classic discussion of robustness in science has an epistemological slant: a scientific result or finding is robust to the extent that it is derivable in multiple and independent ways. Here I come at the problematic from an ontological angle (though epistemology will come into the story too), emphasising a certain robustness that one can associate with the materiality of scientific culture.

To set up the problematic of this essay I find it helpful to think about two stock images of science. One derives from a common-sense view: scientific knowledge is true knowledge of how the world is. On this view, science is not just robust; it is as solid as a rock, given by the world itself. Importantly, it is absolutely *other* to its producers and users. We humans do not have any choice in the matter: the acceleration due to gravity just is 32 feet per second squared. The other stock image of science is the inverse of this. This is the idea of science as a 'mere social construction'—something put together by human beings to suit their interests or to fit in with their social structure or whatever. Here science appears, not as rock-solid, but as extremely soggy, as if any form of knowledge can be projected onto an indifferent and unresisting world. As Barry Barnes (1994) remarks, the world doesn't care what we say about it, so we can say whatever we like. On this view, the otherness of science vanishes. All of the responsibility for specific knowledge claims rests on its human producers, and none on the world itself.

The tension set up by these two mirror-images of science creates the need for a concept like robustness. It is very difficult to put all the weight on the world in accounting for scientific beliefs. Empirical studies seem to point relentlessly to the conclusion that science really is a social construction. The only question that is left is whether it's *merely* a social construction (Pickering 1990). And I take it that speaking of robustness is, then, a way of trying to articulate a sense in which the

A. Pickering (✉)
Department of Sociology and Philosophy, University of Exeter, Exeter EX4 4RJ, UK

Department of Sociology, Kyung Hee University, Seoul, Korea
e-mail: a.r.pickering@exeter.ac.uk

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_13,

'mere' disappears, of trying to get at the idea that the world really can take some of the credit for scientific beliefs, even while acknowledging that they are socially constructed.

That is the line I want to take here. Drawing primarily on the studies and ideas that I set out in *The Mangle of Practice* (1995a), I seek to clarify my own ontological way of grasping the robustness of science, and then ask how this speaks to some traditional philosophical problematics.

The practice turn in science studies (Pickering 1992) has proved destabilising for the philosophy of science, beginning a slide from epistemology to ontology. If the traditional philosophical concern is with what scientists know, the practice turn encourages us to pay attention to what scientists do, and it turns out that one thing that scientists do is to pay close attention to what the world does. So we move from an interest in the doings of scientists to an interest in how nature itself performs and in the coupling of the two. This interest in material performance and agency is by no means a central concern of mainstream English-speaking philosophy of science, but I am convinced it is the place to start in thinking about the robustness of science. If there is a certain nonhuman toughness about scientific knowledge, it is grounded in performative (not cognitive) relations with the material world. That is what I want to discuss first.

Just how do scientists intersect with the material world? Very often not directly with their objects of study; much more often with machines and instruments that generate data for downstream processing. So what does the interaction with machines and instruments look like?

In *The Mangle*, I argued that it takes the form of a *dance of agency* between the human and the nonhuman. In their research, scientists seem to oscillate between bursts of what Ludwik Fleck (1979) called phases of *activity* and *passivity*. In the active phase, scientists are genuine agents, setting up their apparatus this way or that. In the passive phase, they stand back and see what happens. And we can symmetrise the picture by saying that in the phase of human passivity nature is itself active, a genuine agent, doing whatever it will, quite independently of human goals and desires. And then the human agents resume the active role, reconfiguring their apparatus in the light of what they have just found out about how it performs. And then they stand back again and nature resumes the active role, and so on, back and forth—this is what I call the dance of agency. In Chapter 2 of *The Mangle* I dissected this process as best I could in the history of the bubble chamber as an instrument for detecting elementary particles. En route to his Nobel prize, Donald Glaser experimented with all sorts of material systems, putting them together and then literally standing back, with a movie camera in his hand, to record their performance; and then he redesigned and reconceptualised them in the light of that, and tried again to see what the new version would do, and so on.

Over the course of many iterations of this back and forth process, the material form and the material performance of his chambers changed beyond recognition—they were mangled, as I put it—and eventually Glaser arrived at a new instrument, the bubble chamber, that was indeed extremely useful in experimental physics.

So, this is all very simple, but still I think this sort of dance of agency is what we need to focus on if we want to appreciate the robustness of science, so let me continue the analysis. The first point to make is that the bubble chamber was undeniably a human construction. Glaser's active agency was constitutive of its production: he imagined the possibility of constructing a new kind of particle detector, and he put together and reconfigured all of its parts. But we can see at once that the chamber was not a mere social construction. The world may not care what we say about it, but it certainly cares what we do and vice versa. Glaser did not will the chamber into existence; his agency intertwined with material agency in a constitutive way in a dance that he could not control. He had to *find out* what matter will do when arranged this way or that, and this, I think, is the primary sense in which the world enters constitutively into science—and the primary sense in which science is a robust enterprise and not a mere construction.

And I can put this point perhaps more strongly. There was nothing robust about this dance of agency in itself. It was fluid and evolved open-endedly in time. But I think it nevertheless makes sense to speak of the robustness of its product, the bubble chamber. The important thing about the chamber was that it stood apart from Glaser and operated reliably on its own. It was, as I would say, a *free-standing machine* which manifested a sort of *practical duality* of the human and the non-human (Pickering 2009)—it was a material object that acted in the material world quite independently of Glaser or anyone else. Though I need to qualify this idea in a minute, it is worth appreciating the extent to which this does point to a sort of absolute toughness and inhumanity of science, a sense in which science produces and incorporates into itself an utterly inhuman material agency. If the word 'robustness' connects to a feeling that there is something admirable, wonderful, awesome, about science, I personally would locate that feeling in the achievement of such free-standing machines.

So, thinking about the dance of agency and its products quickly and easily gives us an ontological sense of the robustness of science, of how the world itself constitutively enters into science and why science is not a mere social construction. And I therefore want to make a couple of comments on what we have seen so far, before moving on.

One is simply to note that while we have gained an appreciation of the robustness of science I have not yet said anything about scientific knowledge. The dance of agency here was a performative dance not a cognitive one. The bubble chamber itself performed; it was not an idea; it did something in the world. I think philosophy of science, with its epistemological obsession, has spent a long time looking in the wrong place for robustness.

Second, I need to say something more about the otherness of the bubble chamber as a material device. On the one hand, the chamber was a free-standing machine that acted in the world independently of its human users. In that sense it really was other to humanity. On the other hand, we need to remember that the chamber only counted as a successful capture of material agency, as I called it, in a certain social field. At any other time in history it would have looked like a pile of useless junk;

only within a certain configuration of scientific culture did it count as a novel device for detecting elementary particles.

It is at this point that thoughts of *mere* social construction return, as if the state of scientific culture somehow conjured the chamber into existence (which is how Emile Durkheim understood the relation between the social and the technological). To fend off the temptation to think this way, however, we need only to remember that in this instance scientific culture was also reconfigured. In his dance of agency Glaser developed both a new instrument and a form of life that could accommodate it—in accelerator-based physics (rather than cosmic-ray physics, which is where Glaser started off), built and operated by large teams of physicists and engineers, and not by the lone researcher (which Glaser had been when he began his research). So the social did not, so to speak, call all the shots here, and we can hang onto the sense of the otherness of the chamber, as something constitutive of transformations of scientific culture, as something that transmits a certain material otherness to scientific culture, but we should not think of this otherness as absolute—the chamber did not force itself on some passive human world like an alien descending from Mars. We can admire machines and respect their robustness without having to factor out the human side of the dance of agency.

Now I want to widen the discussion. I said that to get hold of the robustness of science one should start with material practice, but that does not mean that scientific knowledge is not important and interesting and that one should not think about it. I have in the past worked through a couple of detailed studies of the production of experimental facts in science, one of Giacomo Morpurgo's quark-search experiments, Chapter 3 of *The Mangle*, the other of the discovery of the weak neutral current (Pickering 1984a, and see Chapter 10, this volume) Both manifest the features I would like to foreground here, but since the quark-search experiments have always been my touchstone for thinking about practice let me talk about them.

The early phases of Morpurgo's experiments were isomorphous with Glaser's. Morpurgo aimed to develop an apparatus that would reliably do something, a free-standing machine—in this case a gadget that would levitate particles of graphite in an electrical field. But the next phase added something new: now he tried to use the apparatus to measure the charges on the graphite particles, looking for the third-integral charges which would signal the presence of isolated quarks. And I find it striking how difficult this simple measurement turned out to be. Again one finds a sequence of active and passive moves in dances of agency, trying this configuration of the apparatus then that, seeing what readings came out of them, reconfiguring the apparatus and thinking about it again, and so on. With one early configuration, Morpurgo indeed found evidence for free quarks. Then he widened the separation of the plates that set up the electric field and the evidence went away. Then he redid the calculations in simple electrostatics that had suggested that the plates should be close together and concluded instead that they should be far apart. This was his first achievement of interactive stabilisation, a point at which the dance of agency extinguished itself in terms of a now precisely tuned machine plus a precisely tuned set of interpretive resources, a place at which his practice could rest and he could publish some results, an articulated fact—namely, the absence of free quarks on some specified amount of matter.

How should we think about this episode, which I take to be typical of empirical knowledge-production in science? The first point might be that any sense of robustness can easily vanish here. Certainly Morpurgo, like Glaser, had built a free-standing machine that performed on its own in the generation of facts. Matter got into the story that way. But the performance of this machine and the conceptual interpretations that Morpurgo wove around it appear to be tied together in a damagingly circular fashion. That this configuration of the machine rather than that was the right one could only be argued on the basis of an interpretive model of the machine, but the rightness of that model was not self-evident and was only guaranteed by the fact that the results obtained fitted in with yet another theoretical model, concerning the presence or absence of free quarks. It is important to see the force of this argument, I think, and it seems to point us back towards an understanding of scientific knowledge as a mere social construct.

How can we escape from this line of thought? As follows, I think. I once used a concept of 'plasticity' to analyse Morpurgo's practice (Pickering 1989), a concept picked up by Ian Hacking (1992). The idea was that the material form of the apparatus and the conceptual form of Morpurgo's understanding of it were not fixed—that they could be bent around and changed open-endedly until they somehow went together and reinforced one another. The trouble with this concept is that it goes very nicely with ideas of the circularity of knowledge production. All that scientists have to do, on this view, is mould the different elements of scientific culture so they fit together. What one cannot get at with this talk of plasticity is why the production of scientific facts is a difficult and uncertain business that can easily fail. And the question thus becomes: where does the plasticity metaphor go wrong?

The point to note is that, while scientists can certainly assemble cultural resources however they like, they cannot know how they will then perform. This gets us back to the dance of agency in an extended sense. Just as Morpurgo, like Glaser, could not know in advance how his material apparatus would perform when configured this way or that, nor could he know in advance where certain theoretical assumptions and approximations would lead him. It just turned out that when he started with one set of plausible assumptions in electrostatics he was led to conclude that the metal plates in his apparatus should be as close together as possible, and that with some modified assumptions he was led to conclude the opposite. So at the conceptual as well as the material level the plasticity metaphor fails, precisely in that while scientists can tinker with their resources however they like, they have genuinely to find out what the upshot of that will be.

So what we have in this instance is not mere social construction but a set of coupled findings-out—finding out where some theoretical calculations will lead, and finding out how an instrument will perform. Again we have to note that Morpurgo was not in control of either of these processes of finding out, nor of whether their products would fit together and *interactively stabilise* one another, as I called it. This was a chancy process, that could have failed. It is a non-trivial historical fact that the performance of the material instrument eventually hung together with one of Morpurgo's theoretical estimates; it did not have to turn out that way at all; they might not have done so; the experiment could have turned out quite differently, both materially and conceptually.

And here again, then, we can salvage a sense of the robustness of science, now of scientific knowledge. Scientific knowledge is not a mere construction, projected onto a passive nature by scientists. The material performance of instruments is indeed constitutive of the knowledge they produce, though prior scientific conceptualisations of the world are constitutive too, and this in an irrevocably intertwined fashion. At the same time, as I said before, this sense of robustness is not one of unsituated otherness. Our knowledge is *our* knowledge, conditioned by the culture it is made from, as we can see in this example, not something forced upon us by nature itself.

I have probably said enough about my overall picture of the robustness of science and how it arises in dances of agency, but I just want to add that I think the picture sketched out so far can be readily extended in all sorts of directions. I think, for example, that one can find dances of agency and interactive stabilisations in purely conceptual practice as well as in the more material strata of science. That was what I just suggested in connection with Morpurgo's interpretive models of his apparatus, and I argued it at length in Chapter 4 of *The Mangle*, taking as my example William Rowan Hamilton's 19th century construction of the mathematical system of quaternions. There I introduced a concept of *disciplinary agency* as a way of talking about what carried Hamilton along to unpredictable places in his development of various algebraic and geometrical formulations, and hence as a way of getting at the non-triviality, the robustness, of interactive stabilisations in purely conceptual systems. I also argued in *The Mangle* that the overall form of my analysis was *scale-invariant*, and that dances of agency punctuated by moments of interactive stabilisation can be found on the macro-historical scale as well as the micro—a claim I tried to exemplify in two later case studies: one of the history of organic chemistry and the synthetic dye industry in the 19th century, the other of coupled transformations of science, technology, society and warfare in and after WWII (Pickering 1995b, 2005). My argument remains, then, that the mangle is a sort of *theory of everything*—though not of the reductive sort beloved of particle theorists.

Now I want to examine the mangle and its associated conception of robustness from some more angles and in relation to various philosophical problematics.

We could start with the problematic of realism. As I said at the beginning, correspondence realism is, I suppose, the starting point for all this talk about robustness. The truth of nature, if scientists had it, would be the ultimate form of robustness. It might therefore be worthwhile examining further just where my analysis departs from realism.

The most obvious departure is that my story of robustness is, as I said before, not in the first instance about knowledge at all. It is about performance in the material world, about what I once called science's *machinic grip* on the world (Pickering 1995a). I find the endless proliferation of free-standing machines and instruments in science enormously impressive—that is where I look for an explanation of the feelings of robustness that science inspires; that is where the otherness of the world enters into science. We can, as I just indicated, draw articulated knowledge into the same picture, by recognising that knowledge production depends on achieving chancy and highly non-trivial alignments and interactive stabilisations of conceptual structures and material performances. One might, then, try to argue that such alignments point to correspondences between the knowledge produced and its objects in the world, but I cannot think why they should, and what interests me most is that my account somehow *defangs* realism and makes it a less pressing topic. If one has no other account of the robustness of science then realism seems very important, the only way to underwrite our intuitions about the otherness of science. But if one has an account like the one offered here, then maybe we could just forget about the entire topic of correspondence: who needs it? We can see and talk about the fact that science is an immensely formidable edifice, by no means a mere construct, without this implausible manoeuvre of picking on one bit of scientific culture—knowledge—and trying to persuade ourselves that it corresponds to the hidden order of the world.

And we could go further with this line of thought. Realism depends on an intuition of *uniqueness*: the world just is one way or the other; science either gets it right or wrong; it would be madness to say that science just gets it wrong; therefore we have to be realists. Ian Hacking, in his book, *The Social Construction of What?* (1999), got at this idea with his conception of 'sticking points'—the points at which scientists resist any kind of constructivist argument. As an example of such a sticking point, Hacking mentions Maxwell's laws of electromagnetism. He himself seems to think that if something like physics were to flourish anywhere in the universe it would eventually have to articulate something like Maxwell's laws; those laws would just impose themselves on the scientists; they are absolutely other to us.

How can I respond to this? I am inclined simply to disagree with Hacking and the physicists for whom he speaks, but I cannot see any way to settle the matter directly. What I can do is discuss the ontological visions that divide us. Hacking's sticking points make sense in a world that really is structured more or less as physicists now describe it. And if it is that sort of place, then probably we were indeed doomed to arrive at Maxwell's laws whether we liked it or not. But everything I have learned from looking into the history of science speaks to me of a world that is not that sort of place at all. It speaks to me of a world that is endlessly rich in endless ways, that can always surprise us in its performance. It is indeed a non-trivial fact that science can latch onto the world in the construction of free-standing machines but, as I have tried to argue, what counts as a free-standing machine depends on who we are and what sort of things we want to do. Within the culture of 1950s particle physics the world revealed itself to us in the shape of the bubble chamber. In the culture of the 1960s, it revealed itself to Morpurgo as having no free quarks. But I find it easy to imagine that different cultures could have elicited quite different machines and instruments and material performances from the world; and I can see no reason not to imagine that. Hacking's sticking points, from this perspective, are facts not about the material world but about the scientific imagination, and its inability to recognise the richness of the world that the history of science itself displays for us.

Another way to put the same point is to note that the analysis of practice I set out in *The Mangle* is an *evolutionary* one, precisely analogous to the evolution of species in a responsive environment. It is a story of continual open-ended searches through spaces of material, disciplinary and social agency and performance, in which specific historical trajectories are marked out by contingent and emergent productive alignments between these elements. This means that transformations of scientific culture in time are *path-dependent*; starting points matter and so do the contingencies that happen along the way. Different starting points and different contingencies should thus be expected to lead to different futures. Stephen J Gould's (1989) vision of biological evolution was that if we could rewind the clock and start the process again then the course of subsequent evolution would be different, leading to a quite different biological world from the one we have now. I think just the same could be said about the history of science. And just as there are no sticking points in biological evolution, I think there are none in the history and future of science.

This is, of course, to juxtapose two visions of what the world is like, and not, as I said, to settle the matter. But I could try to throw a bit more weight on my side of the scales, which will at least widen the field of discussion. I do not know much about the history of Maxwell's theory of electromagnetism, so I cannot argue about that, but I do know about the history of quark-search experiments, and what I know about them increases my aversion to the intuitions of uniqueness that underlie realist philosophy. I think here of the following.

I have discussed my analysis of Morpurgo's quark-search experiments as an explication of the robustness of scientific knowledge in its difficult and chancy alignment with the performance of machines and instruments. But in this case at least robustness evidently did not imply uniqueness. Over just the same period when Morpurgo was reporting his inability to find any evidence for the existence of free quarks on ever increasing quantities of matter, William Fairbank at Stanford was reporting that he could indeed find such evidence. Using apparatus that differed only in specifics from Morpurgo, Fairbank reported several findings of the fractional charges that pointed to free quarks, and a debate grew that encompassed more and more people.

What should we make of this? Two points strike me. The first is that Fairbank achieved just as much of a machinic grip on the world as Morpurgo. Everything I said about Morpurgo's experiments I could have said about Fairbank's. Fairbank's knowledge claims were just as robust in this sense as Morpurgo's. And the conclusion I draw from this is that one should not go overboard about robustness. There is something tough and admirable about articulated scientific knowledge—it is not a mere construction—but this controversy reminds us that it remains situated, relative to particular configurations of material and conceptual resources (Chapter 6 of Pickering 1995a).

Second, we can note that this controversy was eventually settled in practice: Morpurgo's results were taken to be true and Fairbank's false. But the question remains of how this settlement was achieved. Like all of the controversies I have examined in the history of science, there was, in fact, nothing striking or decisive about its ending. One can speak here of yet more manglings of the material, the conceptual and the social, but nothing qualitatively new emerged. Perhaps the best one can say is that the robustness of Morpurgo's results was socially amplified here, as an increasing number of physicists became involved and failed to find either more evidence of free quarks or any productive way of reconciling Fairbanks' results with his. But no-one, I think, claimed that this episode put Fairbanks' claims definitively to rest. At the most empirical level, a great deal of scrutiny of the material and conceptual bases of Morpurgo's and Fairbank's claims turned out not to lead to any solid ground, but came to hinge on details of the electrical behaviour of metal plates they used, something that no-one knew much about. I am inclined to say that the two sets of experiments were *incommensurable* in Kuhn (1970) and Feyerabend's (1975) sense: it proved impossible to find a common measure against which to adjudicate between them.[1]

And to conclude, I want to mention another example of incommensurability, now at the macro-rather than the micro-level. In my book *Constructing Quarks* (1984b), I claimed that one can find two incommensurable regimes in the history of particle physics which I called the old and the new physics. Without going into detail, these understood the world of elementary particles in two very different ways, in terms of very different theoretical models with little overlap. The old physics spoke in terms of constituent quarks and Regge poles; the new physics still spoke of quarks, but now as field theoretic entities, accompanied by a host of other such entities such as gluons and intermediate vector bosons. And the point of calling these historical formations incommensurable, as far as I was concerned, was that they latched onto the material world in different ways. They spoke to different fields of data that were generated by different fields of machines and instruments (colliders instead of accelerators, detectors tuned to 'hard scattering' phenomena rather than 'soft,' computer algorithms that further filtered the data one way or the other).

This macro-incommensurability, then, consisted in almost disjoint machinic grips on the world: the data produced by the machines and instruments of the old physics had almost no bearing on the theoretical concerns of the new physics and vice versa. As Kuhn (1970) put it, the old and the new physics *lived in different worlds*, here in a very down to earth sense. Both, I would say, were admirably robust in the terms sketched out here. There is no suggestion that either was a mere construct. And both, in this macro-example, were able to sustain the practice of large numbers of physicists—so simple social multiplication is clearly not decisive as far

[1] Franklin (1986) notes that at a late stage in the controversy, Luis Alvarez at Berkeley proposed that Fairbank carry out a blind version of his experiment, which is said to have produced random results. With the exception of one mention in a PhD dissertation this story made no appearance in the scientific literature and was not the subject of any technical discussion, so I have not pursued it further. Clearly Fairbank himself did not regard it as definitively setttling the matter: as the *New York Times* reported in an obituary on 3 October 1989: 'Although Dr. Fairbank retired two years ago as physics professor at Stanford University, he had been at work there the night before his death, trying to verify his report of 11 years ago concerning the existence of individual subatomic particles called quarks' (Sullivan 1989). For an account of a very public attempt by members of Alvarez' group to end another scientific controversy at much the same time, by discrediting a Berkeley colleague, see Pickering (1981). The leader of this attempt got so carried away that he accidentally destroyed the nuclear emulsion which was the key piece of evidence in a claim to have discovered a magnetic monpole.

as robustness is concerned. It is, of course, the case that the new physics largely displaced the old in the course of the 1970s, but again I could not find anything philosophically decisive about this process. In the end, the old physics remained viable but was starved of data as programming committees and politicians put their resources increasingly into the hardware and software of the new physics, leading to the overgrown big science we have today.

In summary: I think we can indeed specify the source of science's robustness in dances of agency, especially with the material world, and in the production of free-standing machines and instruments, but we should not overrate this robustness. Machines, instruments and knowledge are machines, instruments and knowledge only in relation to us, situated in a path-dependent fashion with respect to the cultural fields in which they are built, fields which themselves co-evolve in a chancy fashion with those machines, instruments and bodies of knowledge. To put robustness in its place, and to undermine the intuition of uniqueness that goes along with our taken-for-granted realism about science, I have tried to suggest that incommensurability is always bubbling up in science, at all scales from the micro to the macro, and that science has no magic recipe for getting round that—only more manglings and dances of agency.

Traditionally, philosophy of science has done all it can to reject, or simply ignore, the very idea of incommensurabilty. Personally, I think it makes philosophical thought much more interesting. One has to learn to imagine an indefinitely rich world which we can latch onto in an indefinite number of ways. I think we can do that, and still appreciate the robustness of science at the same time.

Acknowledgement This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-330-B00169).

References

Barnes, B. 1994. "How Not to do the Sociology of Knowledge." In *Rethinking Objectivity*, edited by A. Megill, 21–35. Durham, NC: Duke University Press.

Feyerabend, P.K. 1975. *Against Method*. London: New Left Books.

Fleck, L. 1979. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Franklin, A. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Gould, S. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: Norton.

Hacking, I. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, edited by A. Pickering, 29–64. Chicago: University of Chicago Press.

Hacking, I. 1999. *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Kuhn, T.S. 1970. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.

Pickering, A. 1981. "Constraints on Controversy: The Case of the Magnetic Monopole." *Social Studies of Science* 11(1):63–93.

Pickering, A. 1984a. "Against Putting the Phenomena First: The Discovery of the Weak Neutral Current." *Studies in History and Philosophy of Science* 15:85–117.

Pickering, A. 1984b. *Constructing Quarks: A Sociological History of Particle Physics*. Chicago: University of Chicago Press.

Pickering, A. 1989. "Living in the Material World: On Realism and Experimental Practice." In *The Uses of Experiment: Studies of Experimentation in the Natural Sciences*, edited by D. Gooding, T.J. Pinch and S. Schaffer, 275–97. Cambridge: Cambridge University Press.

Pickering, A. 1990. "Knowledge, Practice and Mere Construction." *Social Studies of Science* 20:682–729.

Pickering, A., ed. 1992. *Science as Practice and Culture*. Chicago: University of Chicago Press.

Pickering, A. 1995a. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Pickering, A. 1995b. "Cyborg History and the World War II Regime." *Perspectives on Science* 3:1–48.

Pickering, A. 2005. "Decentring Sociology: Synthetic Dyes and Social Theory." *Perspectives on Science* 13:352–405.

Pickering, A. 2009. "The Politics of Theory: Producing Another World, with some thoughts on Latour." *Journal of Cultural Economy* 2:197–212.

Sullivan, W. 1989. "Prof. William N. Fairbank, 72, Physicist and Pioneer in Quarks." *New York Times*, 3 October 1989.

Wimsatt, W. 1981. "Robustness. Reliability and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M. Brewer and B. Collins, 124–63. San Francisco, CA: Jossey-Bass.

Chapter 14
Dynamic Robustness and Design in Nature and Artifact

Thomas Nickles

14.1 Introduction

My chapter applies a bit of qualitative risk analysis to systems of inquiry and their products. It extends Charles Perrow's theses about "normal accidents" in technological systems to epistemic systems, that is, to humanly constructed (explicitly or implicitly designed or engineered), evolved and evolving complex technological systems of *inquiry* and their products. My focus is on enterprises of ongoing scientific research at innovative frontiers. My central claims are: (1) Although the robustness (in Wimsatt's sense) of a scientific research program and/or its products is obviously highly desirable, no improvements in robustness can render these processes or their products invulnerable to failure. (2) On the contrary, such improvements can often, as far as we know, make inquiry systems vulnerable to new kinds of failure, sometimes worse failures than before. Robustness in one dimension can render a system more vulnerable to catastrophic change in another dimension. (3) Thus robustness is a relative rather than an absolute concept. Rather than vanquishing fragility, complex robustness can shift its location. More than that, increasing robustness (e.g., by adding new experimental or conceptual linkages) can actually generate fragility where none existed before. Accordingly, we cannot expect to make uniformly cumulative progress toward risk reduction. My argument can be construed as a (relatively new?) attack on foundationism and also on strong forms of convergent epistemic realism in the sciences. (4) In evolving epistemic systems with lookahead, *prospective* robustness is crucial to decision-making, contrary to traditional, retrospective empiricist theories of confirmation and robustness. I shall employ a broadly Kuhnian conception as my representation of mature science, that is, science with well-established problem-identifying and problem-solving routines. Any particular choice is, of course, controversial as well as arbitrary, but Kuhn's model is the best known.

T. Nickles (✉)
Department of Philosophy, University of Nevada, Reno, NV, USA
e-mail: nickles@unr.edu

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5_14,

14.2 The Robustness-Fragility Tradeoff

In recent years the term ‘robustness’, its cognates and neighbors (solidity, persistence, hardiness, reliability, resilience, viability, flexibility, healthiness, etc.) have been applied to just about everything. In fact, ‘robust’ has become a buzzword in popular culture that can be applied to anything that exhibits strength of some sort. We want our aircraft to be as resilient as possible to turbulence and to various types of mechanical failure. We want our electrical grids to be robust to local failure, to unusually high power demands, and to insults such as severe storms. We want our nuclear plants to have monitoring and backup systems reliable enough to prevent the expected occasional malfunctions from leading to catastrophic accidents. We want our scientific research programs and experimental systems to be sensitive to empirical signals yet tolerant of modeling flexibility (abstraction, idealization, simplification, and approximation) and even errors of certain kinds. And so on.

We have made a lot of progress on all of these fronts. Thus the following question arises. While perfection is unattainable, can we not expect to approach it as an ideal, so that the only failures will be relatively small ones? Will not our progress toward greater robustness be cumulative as we successively minimize existing sources of error? Will not our successful error-reduction efforts converge on perfection in the limit? Let us term this *the cumulative fragility-reduction thesis* or *convergent risk-reduction thesis*, labels that are clumsy but descriptive. Surely the thesis is true and thus identifies a realizable methodological goal?

No, said organizational sociologist Charles Perrow in *Normal Accidents* (1984), not for technological systems (cf. Gertstein 2008). Perrow argued that the components of tightly coupled, complex technological systems will normally experience unexpected, untested, and practically untestable interaction effects. Such “interactive complexity,” as he called it, becomes apparent in accidents involving multiple failures, accidents such as the Apollo 13 space-module problems of 1970 (problems that, fortunately, were handled so as to avoid disaster); the DC-10 air crashes, including the Turkish Airlines plane over Paris in 1974; the partial meltdown of the nuclear reactor at Three Mile Island, Pennsylvania, in 1979; and the disastrous explosion at Chernobyl in 1986. A more recent example is Hurricane Katrina’s devastation of New Orleans in August 2005. In this case the very water control systems previously constructed by the U.S. Army Corps of Engineers and other agencies significantly worsened the failure by channeling the water in a destructive way. Still more recently, it is likely that the spring 2011 earthquake, tsunami, and nuclear disaster in Japan will be another case, once the details come to light. The deep recession of 2008–2011 was a failure of internationally linked economic systems apparently triggered by Wall Street’s opaque bundling of risky financial derivatives as well as governmental regulative laxity. Mutual defense treaties have similar features. The very attempt to improve a nation’s security can make it vulnerable via an attack on one of its friends.

Perrow’s second major contention is that these “accidents” should be considered a *normal* part of the operation of such systems rather than as highly contingent insults from outside the system. They are endogenous, not exogenous. They are intrinsic to the system and are to be expected, in a generic sense, although not, of course, specifically. (A hybrid case is accidents triggered by an external event, in which the response makes the situation worse.) Third, the operators of such systems should not be saddled automatically with the blame for failure, even if a precise sequence of procedures could have saved the day; for such failures are extremely confusing. Human beings are not omniscient and, typically, the event cascades are rapid and the incoming data and advice that the operators do get (as given by meter readings, indicator lights, warning horns, from experts employed by the manufacturers, etc.) are conflicting or otherwise unreliable. Failures of this sort, Perrow says, should be regarded as system failures, not operator errors.[1]

Perrow's fourth main point is that adding further safeguards against such errors only adds to the complexity and hence introduces new sorts of fragility into the system, even as it increases robustness elsewhere. For such additions typically increase exponentially the number of possible interaction effects, which, by nature are nonlinear.

Perrow's ironic conclusion is that there is a direct coupling of robustness to fragility. An increase in robustness does not mean an absolute decrease in fragility. Striving for greater robustness increases complexity that, in turn, commonly creates new paths for potential failure, including major malfunctions. Thus even successfully reducing the malfunction rate from expected sources does not result in a cumulative gain, as far as we can tell. There is a tradeoff. The very effort to eliminate fragility and catastrophic failure is, to a degree, self-undermining.

My purpose in this chapter is to extend Perrow's insight to epistemic systems, to systems of inquiry and their products. For they, too, can be regarded as complex technological systems. A theory or model can be regarded as a design, but so can a research program. A field can be represented by linked networks of several kinds involving personnel, equipment, "natural" materials, research designs, social support and demand systems, published papers, and the like. Or so I shall assume without argument.[2] I shall also assume that Wimsatt's insightful analysis captures a broad sense of robustness in the sciences, at least at a relatively high level of description (Wimsatt 1981, 2007, Chapter 7; see Chapter 10). The sense of robustness with which I shall be most concerned is one in which a system is sufficiently responsive to a variety of internal and external shocks that may befall it as to maintain its viability. This conception overlaps Wimsatt's in the sense that both involve a kind of invariance under changing circumstances.

[1] See also Perrow (1972, 2011). It is also relevant to mention here Herbert Simon's seminal work on bounded rationality and the behavior of complex organizations (Simon 1947). Bill Wimsatt and I share an admiration for Simon's work. As I once told Bill, I regard Simon as one of the great American pragmatists.

[2] Although I do not have space to defend this claim, most of the people represented in this volume surely accept some version of it. Incidentally, while large portions of the design are deliberately engineered in response to inputs from "nature," as is usual in cases of large human constructions, we should expect that modern sciences and their research programs contain important elements that were not explicitly designed, some of which we are surely unaware.

My central claim (combining claims 1 and 2 of my Introduction) is that the convergent risk-reduction thesis is false when applied to epistemic systems. The very attempt to control certain kinds of risk or error can generate new sources of risk that are rarely realized but that are generally more difficult to predict and to handle when they do. Robustness is engaged in an intricate dance with fragility. The implication for us is a heightened sense of the fallibility of our research systems and their products, an implication that makes my position on issues such as epistemological realism broadly compatible with those of Pickering and Soler (Pickering 1980, 1995; Chapter 10). I certainly do not deny that there has been rapid scientific progress on many fronts. However, I am not confident that even our most mature sciences tell us which entities and processes *really* populate our universe at bottom. A robust scientific realism that tells us what the universe is really like (thereby replacing the metaphysics of yore) is not at hand, especially when it comes to the very large and the very small. And the complexity-theory worries that I sketch here raise difficulties even for theories and models "of the middle range."

How is it possible that the risk-reduction thesis is false? How can increasing robustness at the same time decrease robustness? The obvious answer is that we must relativize robustness to specific dimensions or types of failure. And this is my third thesis. Robustness is not simply a matter of degree, as it is sometimes depicted). It is *relative* to specific kinds of external insults and perturbations and to internal breakdowns or constructive changes as well. Robustness is a relational property, represented by a two-place logical relation at a minimum and better as a three- or four-place relation. We can say that system s is robust to perturbation p to degree d except when c. In symbols, $R(s,p,d,c)$. The last clause may be optional, considered to fall under the usual ceteris paribus clause. However, operating manuals often provide explicit exceptions that seem stronger than ceteris paribus clauses.

I cannot of course prove that these theses hold in every case. Indeed, I don't think that they do. But I claim that they do hold for important kinds of epistemic systems, especially those such as we find in the sciences that place a premium on innovation and which, as a result, experience dynamic change over time. That is enough, I believe, to challenge the cumulativity thesis.

Whereas Perrow considered accidents in relatively static systems, once they are designed and built, we must consider also the crucially important case of systems that are designed to be dynamic in a strong sense—that contain intrinsic, endogenous sources structural change. How is it possible that such systems can be viable, that they can robustly maintain functionality throughout the change? For living systems, this question has both an ontogenetic and a phylogenetic form. How is it possible that biological organisms are robust to sometimes extreme developmental changes, as in the life-cycle of a butterfly? And how is it possible that genetic and other systems continue to operate reliably through sexual-recombinant and mutational change? Over time our human ancestors changed from small mammals to large, upright creatures, not to mention the earlier lives of plants and, long before that, single prokaryotic cells. In this sense we still face a version of the fixity-of-species problem: How is it possible that species can evolve? As Wagner (2005) points out, changing the genome (including patterns of gene regulation) changes the basic operating instructions for organisms, and these changes are heritable by offspring; so the fate not only of a particular individual but also that of an entire species hangs in the balance.[3]

The problem, then, is to understand how biological nature or intelligent beings can design a system robust to *internal* design changes, to technological upgrades, so to speak, occasionally even in deeply *entrenched* subsystems, without loss of functionality (Wimsatt 1981), and to appreciate the risk of catastrophic failure in such systems. Once we reject the foundationist impulse, we realize that epistemic research systems are strongly dynamic in this sense. Presumably, this holds for any enterprise that places a premium on suitably adaptive, creative innovation. As inquiry proceeds, even the deepest principles can be overturned and the enterprise restructured. Historically, many such enterprises have failed as a result, but others have survived major transformations. The Internet is a recent technological example of the latter sort (Willinger and Doyle 2005; Doyle et al. 2005).

Since the 1960s, the debate about scientific revolutions has turned partly on these issues. Those sympathetic to Thomas Kuhn's view in *The Structure of Scientific Revolutions* (1962, 1970) regard scientific revolutions as radical events akin to those political revolutions, such as the French Revolution, that throw out the old social order and send the community off in an unexpectedly different direction. These analysts are in turn challenged by those who stress continuity rather than failure. In any case, as Kuhn himself emphasized, scientific revolutions are extremely creative episodes in which the old paradigm is regarded not as a complete failure that ends the enterprise but, rather, as a failure only relative to a promising new approach that somewhat reinvents the enterprise, returning it to robustness as a progressive site of ongoing research. In the words of the economist Schumpeter (1942), we can identify "waves of creative destruction" in the history of science as well as in economic and technological history. Here we can distinguish enterprises that are truly left behind from those that reinvent themselves, maintaining a continuity of some sort. In the sciences, perhaps more than in technology and the general economy, the transformation is often enterprise-preserving creative destruction. I return to Kuhn's work in Section 14.9.

The topic of strongly dynamic epistemic systems leads me to my fourth thesis. Unlike the rest of nature, the human designers of epistemic systems possess a degree of lookahead. We humans can think about future prospects and consequences, make plans, and, to some degree, shape our own opportunities. Scientists in particular are capable of assessing the future promise of various alternatives and of making corresponding decisions today and then proceeding to reprogram themselves, so to speak. For systems involving creative inquiry, then, I want to suggest that there is a

[3] Wagner and co-researchers are engaged in an extensive research program to explore these questions involving robustness, embryological development, evolution, specialization, modularity, and the like. For example, Martin and Wagner (2008) discuss the tradeoffs in genetic networks that serve more than one function. They ask to what extent the need to serve several functions constrains the network architecture—and what effect such compromises may have on robustness.

prospective dimension to robustness, one sometimes so powerful as to trump *retrospective* judgments of robustness based on empirical track records such as reliability tests and multiple derivations of a result (Wimsatt 1981).

As a designer, biological evolution has major advantages over us in some respects. For example, almost every single organism (of the zillions on earth, past and present) "field tests" a distinct design variation; and nature is not constrained by our human horizons of imagination. But we have some advantages as well. For example, biological nature is more limited than human designers are in the respect that humans can often start over from scratch, undertake fundamental revisions, and combine previously distinct technologies. Furthermore, many of our artifacts do not require continuous functionality: we can tinker with them in the laboratory without releasing them into the field. The epistemic systems of the sort that I ultimately want to consider as part of a larger project are those that have managed to preserve functionality under transformative change, sometimes on the basis of lookahead. That is why I speak of *inquiring* systems: to bring out the fact that these are dynamic enterprises. The most important part of epistemology, in my opinion, is "frontier epistemology," the study of how knowledge practices grow at the frontiers of research and how the various modes of inquiry often manage to survive the surprises encountered there.

Perrow's claim about normal accidents is interesting for other reasons, for a normal accident is, somehow, less contingent, less accidental, than an accident as commonly understood. This point about relative contingency can be generalized and has important implications for several important problems, especially those concerning dynamical change in a system. For example, I believe that Perrow's insight concerning normal accidents as intrinsic to complex systems has a happier bearing on the problem of endogenous innovation. Accidents are usually considered exogenous—system insults from outside, mere contingencies rather than *systemic* features. Whereas for Perrow most accidents are normal, part of the behavior arising from within such systems, part of the normal background noise of complex systems. Given the extreme nonlinearity of such systems, an ordinary event such as a valve getting stuck or accidentally being left closed after testing can cascade into the meltdown of a nuclear reactor.

I extend this insight toward an endogenous account of *innovation*, which is also highly nonlinear and unpredictable, but dependent on a sort of ordinary background noise—that is, variation—intrinsic to the system. In some cases, a relatively ordinary or "normal" sort of innovation can have revolutionary consequences, producing, now, an exciting cascade of problem *solutions* or technical *innovations* rather than disaster. Here we have returned to the problem of Kuhnian revolutions. Members of the old guard may indeed consider a revolution a disaster, as destruction, while those in the vanguard see it as a creative continuation of the overall enterprise. There can, of course, be breakthroughs that are not destructive of the received platform.

14.3 Tradeoffs Between Robustness and Fragility: A Variety of Examples

First a political example, the reign of Louis XIV (d. 1715), who supposedly asserted *L'état c'est moi* and insisted that a united, coherent nation-state must have *un roi, une loi, une foi*. At the time the nation-state was a relatively new entity and violent power struggles were common. (In his social contract theory of the state published a few decades before, Hobbes had noted that a single powerful ruler was a stabler arrangement than a ruling body of two or more persons.) Under these circumstances, the rigid hierarchy that Louis imposed created a robust system that minimized the danger of both external and internal threats via a triple structure of military, legal, and religious power. And yet that very hierarchy made the system fragile or brittle in another respect. Strongly hierarchical, "command and control," hereditary regimes such as Louis XIV's are robust to confusions about chain-of-command; but they are fragile to questions of succession and to failure at the top, for there is nothing in the system to guarantee a strong leader. This turned out to be the case in French history. After Louis' death, the succeeding Louises were weak and vacillating. The story is of course very complicated, but the cumulative result of many factors was the French Revolution. Complexity theorists Levin et al. (1998) observe that, still today, rigidity is often taken to be the mark of robustness in social systems. Such systems withstand the forces of change, but this ultimately makes them seem antiquated. When they fall, they tend to fall quickly. The fall of the Soviet empire is a recent example.

As noted, one of Perrow's own primary examples was nuclear power plants. Designing complex safety mechanisms and backup systems certainly improves robustness against anticipated failures, but it creates new routes to unexpected failures, e.g., when a failure of one system masks failures in others, including human error.

The "Six Sigma" (6σ) quality control program in industry aims to reduce variation in manufactured products so that there are fewer than 3.4 failures per million.[4] Here 'robust' production clearly means elimination of unwanted variation, noise, and waste. Later we shall have reason to question the assumption of a normal distribution in some such processes. However, the following illustration has a different slant. Consider the recent change in the 3M Corporation (formerly Minnesota Mining and Manufacturing), makers of many kinds of tape and thousands of other products. 3M has a history of encouraging innovation at all levels, e.g., by building innovation time into many employees' contracts and setting as a company goal that 30% of annual sales will be of products introduced within the past 4 years. The result has been a robust corporation that has flourished by repeatedly reinventing itself for over one hundred years, far outlasting many competitors. In an effort to make the corporation even more robust, its products more reliable, the new CEO of

[4] Actually, this failure rate calculates to 4.5 "sigmas" or standard deviations from the mean value, assuming a normal distribution. The extra 1.5 sigmas are imposed to allow for long-term variation, given that the approach is evidence-based and that most evidence is short-term.

3M recently implemented Six Sigma quality control and a rigid system of employee evaluation. Unfortunately for 3M, there has been a tradeoff: less time and incentive for innovative *intra*preneurship (Hindo 2007). Innovation is typically a process that looks sloppy and inefficient from the outside, since research and development depends heavily on variation-selection processes with many false starts. It requires slack. It requires time. The primary cost of new products is in the design stage, not in final production cost per unit. Rigid quality control in the generative process can have serious long-term consequences for innovation.

For a classic epistemological example, consider Descartes. His foundationism had the goal of building knowledge structures that are impervious to future criticism and thus guaranteed to last. The model of a finished system was Euclid's geometry, an axiomatic deductive system.[5] Although such a system is robust to failures of logical entailment and can deductively systematize a remarkable amount of content by means of a few postulates and inference rules, we can identify several sources of fragility or brittleness in such systems—and in Descartes' in particular. (1) His starting principles were not as perfectly robust ("clear and distinct") as he believed. (2) Several of his inferential steps are deductively questionable, and there is the well-known problem of circularity in the early steps of his system. (3) The extreme verticality of the Cartesian foundational hierarchy is itself cause for concern, as Fig. 14.1 immediately suggests. The system teeters in unstable equilibrium.

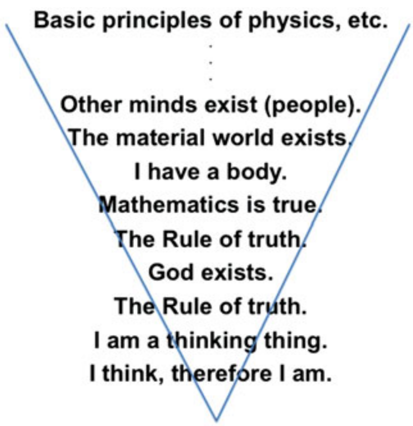

Fig. 14.1 The supposed Archimedean point of Descartes' system is the *cogito ergo sum* at the *bottom*, which allegedly defeats the most severe forms of skepticism. Then come, in order, the other steps of *Discourse on Method* and *Meditations*: I am a thinking thing, the Rule of Truth (Everything that I perceive clearly and distinctly must be true), etc. The entire enterprise is supposed to provide an escape from history (thus from historical path-dependence other than its own, intrinsic, cumulative history of logical development) in the sense of giving us epistemic foundations so robust that they are impervious to the ravages of time and circumstance. (Figure by the author)

[5] Descartes admitted algebraic derivations as well as geometrical ones, and his use of 'derive' was more liberal than that allowed by the later, formal concept of deduction.

(4) Shockingly, failure of types (1) or (2) propagates uncertainty *instantly* through the remainder of the system, even if—or especially if—those couplings are tight, that is, deductively valid. Ironically, the very thing that makes the system epistemologically well founded, highly integrated or systematized, and economical, with instantaneous transmission of derivability from starting points to manifold end points—and thus so robust in these respects—makes it fragile in another respect. "Instantaneous logical action-at-a-distance," as we might call it, can be a very good thing—or a very bad thing. Tradeoffs! The couplings of the propositions are so tight that an error anywhere produces a disastrous cascade of failure through everything else that depends on that step. There is no way to stop it. The system is vulnerable to epidemic contagion of error, or at least uncertainty (which was tantamount to error for Descartes-the-foundationist). However, things are not so bad for methodologists who believe that empirical support comes from consequential testing rather than from antecedent premises. For them strong deductive coupling has the advantage that a predictive failure propagates backward in such a way as to help them to root out fundamental error, modulo Pierre Duhem's fault-assignment problem.[6]

(5) Our confidence in a system that pretends to be failsafe plummets when even a small failure is detected. This is surely one reason why some critics find it (too) easy to discredit modern science. In their view, scientific knowledge claims are supposed to be nearly infallible; so, for them, every reversal of scientific judgment is damning.

The chapters of this volume contain several examples from recent experimental research of control systems for robustness. Soler and other contributors, adding to the growing science studies literature, are advancing our understanding by showing how many, how esoteric, and how intricately coupled are the investigative techniques, regulative procedures, standards, and professional judgments involved in establishing or negotiating an experimental claim. This in contrast to those early philosophers of science who considered "experimental observation" relatively unproblematic, epistemologically uninteresting, and requiring no special attention. Concerning one stage of the Gargamelle search for weak neutral currents in which experts checked the tracks of nearly 300,000 photographs, Soler writes:

> [T]he aim is to extract a set of tracks whose interpretation is globally more reliable, in such a way that it becomes less likely to make mistakes in the counting of the potential tracks of NC [processes identifiable as neutral currents].
>
> But if this is the aim, experimentalists are never sure that it is indeed achieved. The filtering operations aim at eliminating some confusions, but they can themselves be sources of mistakes.

The elaborate control systems for detecting and avoiding error that we build into our experimental technologies open up new routes of potential error.

In epistemology and methodology, philosophical thinkers have tried for centuries, without success, to develop fail-safe systems, systems and methods so robust

[6] I refer to Duhem (1954) on the need for auxiliary assumptions in any predictive inference and the resulting difficulty of pinning the blame for failure on any one premise.

that all threats, all sources of failure, are eliminated.[7] Operationism as a method of concept formation is a recent example.[8] By now, however, most thinkers have declared themselves to be fallibilists of one kind or another, ranging from still-confident, strong epistemological realists to skeptical antirealists. But the strong realists would appear to be weak fallibilists in the sense that they apparently believe that, although we cannot totally eliminate error, we can identify its sources and cumulatively reduce its occurrence without thereby creating new sources of error. It is this cumulativity thesis of error reduction that I am combating, following Perrow and the complexity theorists discussed below. If I am right, then, despite the tremendous advances of recent, mature science, we must remain thoroughly fallibilistic—in part *because of* those very successes! Karl Popper was a thoroughgoing fallibilist (e.g., Popper 1963), but even he based his claims for verisimilitude (approach to the truth) and realism, in part, upon this idea of gradual error elimination.

Wimsatt (1981), citing Feynman (1965), distinguishes linear, "Euclidean" intellectual structures from "Babylonian" structures (see Chapter 8). The latter are multiply connected. In a Babylonian structure, such as a truss bridge, if one link fails, the structure does not collapse. In the intellectual counterpart, if you forget one way of deriving a result, you can resort to other ways, since they are interconnected. Your "failure" is contained. Moreover, unlike the Euclidean intellectual model, the support is mutual to varying degrees. Justification does not flow in one direction only, from foundational axioms to theorems (or inductively, from good test results to hypotheses). As Putnam (1962) observed long ago, it can go in many ways at once.

Thus we come to the idea that there can be types of virtuous circularity as opposed to the vicious circularity inherent in Descartes' system. In logical-semantic structures as in causal structures we can have mutual support. Pickering (1995) develops this idea in terms of symbiosis. Today's network analysts frequently deal with what they sometimes call "circular causality."

---

[7] Here we can include not only philosophical and scientific systems based on reason and evidence but also ethical and religious systems based on revelation, mind-control, and various other self-protecting contagions and social viruses discussed by meme theorists. The most secure, deeply entrenched of these is, according to the line being developed here, vulnerable to catastrophic collapse. Gaye McCollum-Nickles reminds me here of Oliver Wendell Holmes, Jr.'s poem, "The Deacon's Masterpiece or, the Wonderful 'One-hoss Shay': A Logical Story," about the sudden collapse of Calvinism in late 19th-century America. Pickering (Chapter 13) regards standalone, material machines as the hallmark of modernity. Also falling within modernity is the extension of this idea to Bacon's and Descartes' (unsuccessful) attempts to mechanize scientific procedures (the alleged discovery of "the scientific method"). As a response to an unpredictable world, the idea of "standalone" ethical systems invulnerable to the arrows of fortune is very old. Such were the systems of the ancient Stoics and Epicureans.

[8] Bridgman (1927), one of the founders of operationism, blamed the need for the relativity revolution on physicists' failure to operationally define their concepts (especially simultaneity) prior to theorizing—as if we could work out a robust system of concepts prior to theory in work at the frontier of research!

We must be careful, however, for, insofar as a Babylonian *intellectual* structure is tightly coupled by deductive relations (or something similar), error or uncertainty can spread rapidly there as well—in fact "in all directions at once" in which there are tight structural connections. To be sure, scientists possess various devices—buffers, spacing, and security walls—for containing failure in order to avoid devastating cascades. This point has been noted by several authors, e.g., Quine (1951) on the web of belief with its centralities and priorities, Lakatos (1970) on research programs with their "protective belts," and Wimsatt (1981, 2007) on walling off difficulties. But if I am right, nothing that we can humanly do can prevent occasional, surprising avalanches of failure. At the very least, the topic deserves further study.

14.4 Highly Optimized Tolerance (HOT)

Jean Carlson and John Doyle (physical scientists at the University of California, Santa Barbara, and Caltech, respectively) utilize the latest tools of network theory and percolation theory to develop a Perrow-like thesis (Carlson and Doyle 1999, 2002). They speak of "spirals" of complexity. Attempts to improve robustness lead to greater complexity which in turn generates new kinds of fragility, which lead to new security mechanisms, and so on. This amounts to a different sort of "arm's race" than the traditional one between predator and prey, or at least provides a different perspective on the latter as a special case.[9] Carlson and Doyle see the quest for robust viability as the driving cause of complexity. For them pressure for greater robustness *explains* complexity in a sense that generalizes evolutionary biology to include technological design: greater robustness is greater "fitness."

So why not keep it simple? Why start the complexity spiral in the first place? Why not just stick to simple, reliable systems?

Carlson and Doyle make an obvious but useful distinction between two kinds of robustness: (1) simple systems made of a few highly reliable components and (2) complex systems of "sloppy," cheap components, where the robustness is an emergent, systemic feature deriving from backup systems, extensive monitoring, automated, computerized control, and so on. The problem with simple robustness is that it is just too simple. Such systems operate in too narrow a range, with systemic failure looming beyond. Relative to a desired wider range of operation, complex systems can be far more robust. As Carlson and Doyle point out, a Boeing 777 is more robust to variable weather than a two-person airplane with simple instruments. Besides, nothing is failsafe anyway. Even simple systems sometimes fail. When a

[9] Since robustness includes resilience to environmental shocks in addition to predator-prey issues, even in biology, the phenomenon is more general than the usual sort of arms race in which, e.g., long legs for greater speed increase the fragility of the legs. Our primary concern here is with human technological systems, including epistemic systems. We can construe the race as an attempt to identify and avoid possible new sorts of accidents before they happen, or before they happen again.

component of a simple system fails, there can be no cascading failure, since there is no complexity to support a cascade; but the failure is typically disastrous in any case since simple systems are not likely to degrade gracefully. Complex systems with redundancy, networking, and/or high-tech monitoring and feedback control can typically minimize the damage from isolated, random component failure, because they do not demand consistently high tolerance in individual components.

Highly designed systems are more expensive, but often the additional efficiency or yield more than compensates. This is a second important feature of their HOT systems, those manifesting highly optimized tolerance, namely, the pressure for optimization. The physiology of an elephant is far more efficient than that of an ant, and a Boeing 777 is more efficient (in load carried per amount of fuel) than a Piper Cub.[10] And, again, complex systems can employ sloppier, hence cheaper components.

The upshot of all this is that we cannot keep it simple: we must deal with complex systems. The demand for robustness requires complexification of the control structures. The question then becomes which architectures are more robust than others. The general strategy that Carlson and Doyle take is not to avoid all failure, which is impossible, but to prevent *anticipated* types of failure from becoming so highly contagious that an epidemic or failure cascade ensues. And the way to do that is to maintain adequate spacing, metaphorically speaking, around danger areas, to install buffer zones in order to confine the damage. Their favorite illustrative model (a standard one) is forest fires, where the spacing is quite literal. Zones subject to high lightning strike rates or to heavy human (mis)use should have buffer zones around them to keep any fire from spreading very far.

Installing large enough buffer zones will, of course, nearly always prevent devastating fires. However, that recourse also cuts down on timber production (or the amount of forested land preserved for any reason, e.g., as an ecological system). So the problem becomes how to optimize "throughput" or productivity or yield (the harvested wood in their toy model) while maintaining adequate robustness. Maximizing productivity or overall "fitness" is where the 'highly optimized' of their Highly Optimized Tolerance (HOT) research program comes from. The tolerance refers to the robustness.

Carlson and Doyle maintain that maximal productivity does not occur in merely physical systems, for it requires *design*, either biological evolutionary design or deliberate engineering design as in human technology.[11]

[10] See Geoffrey West's lecture, with slides, available at http://online.itp.ucsb.edu/online/pattern_i03/west/. Power laws apparently characterize the metabolism rate, energy use, extinction rates, and many other aspects of the animal and plant worlds. For criticism, see Downs et al. (2008). For a survey of leading models of species extinction, see Newman and Palmer (2003).

[11] In my opinion the distinction between novel "design" by natural selection and intelligent design by human engineering is usually exaggerated. While there are important differences, at bottom both are selectionist processes, that is, variation-plus-selection processes. See Nickles (2003 and forthcoming).

> The central claim from our model is that the essence of this robustness, and hence of biological complexity, is the elaboration of highly structured mechanisms that create barriers to cascading failure events. (Zhou et al. 2002, p. 2053)

They argue that highly specialized design can provide remarkable robustness for anticipated failures but is extremely fragile to *unanticipated* kinds of changes, external and internal, as well as to design errors. For example, biological species that track their environment very closely become "specialists" vulnerable to environmental change, at which point the "generalists" will survive or move in and take over (Calvin 2002).

Carlson and Doyle defend their claims with a technical analysis that I cannot repeat here. A central point is that the "size" of failures in complex systems in HOT states, that is, systems that are highly optimized in the indicated manner, have distributions with "heavy tails" (or "fat tails" or "long tails," as they are also called). Heavy tails of just the right shape are the signature of power laws. Heavy-tailed probability distributions decay at slower rates than the exponential drop-off characteristic of Gaussian normal distributions. The former are sub-exponential over at least part of the tail.

With a Gaussian distribution the three main kinds of averages (mean, median, and mode) coincide, and the rapid drop-off means that very few cases exist that are more than three standard deviations from the average (the more so as the peak is narrower). Thus it makes good sense to speak of a "typical" item or event, namely, one close to average. This determines the "scale" of the phenomenon. The mean together with a well-defined variance or spread of the distribution provide a neat, two-parameter summary of the phenomenon in question.

Not so with heavy-tailed distributions. These are characteristic of scale-free (scale-invariant) phenomena. Here one can get large events as far out on the tails as you please, with non-negligible probability. There is no such thing as a typical size. The variance is no longer well defined; in effect, it becomes infinite, modulo the limits of system size. The bad news is that, when the events in question are failures, failures of any magnitude are realistically possible. There is no preventing occasional disaster. To illustrate the difference: human height is normally distributed. It is extremely improbable to encounter a human who is eight feet tall and virtually impossible to meet one ten feet tall; but huge earthquakes, although rare, are not "virtually impossible." And—to revert to a previous example—insofar as the reliability of industrial processes can be characterized by distributions with heavy tails, it makes no sense to speak of "six sigma" reliability. For heavy-tailed processes, traditional risk analysis underestimates the true danger—the probability that an event with a large negative utility will occur.

In sum, Carlson and Doyle offer a technical understanding of Perrow's compromise. The cost of controlling robustness in such a way as to maximize throughput is susceptibility to unanticipated sorts of disasters. But Carlson and Doyle perhaps go further than Perrow to consider also constructive changes internal to the complex systems in question. The problem of how a system can withstand variability of its components is part of the problem of managing failure of traditional, static systems; but how account for the viability, the flexibility, the adaptive capacity or evolvability, of complex systems that undergo basic design changes that could improve performance? After all, the process of optimization is itself an ongoing process. As noted above, Andreas Wagner confronts this problem for biological evolution. Doyle and Carlson, being physical scientists, take the Internet as their chief example of such a system (Doyle 2005; Willinger and Doyle 2005), showing how the system can be robust to rapid technological changes at the applications level. A similar problem arises for an ongoing scientific research program, e.g., in what Kuhn calls normal science. Philosophers have worried most about continuity through scientific revolution (incommensurability and all that), but for Kuhn it was the strong continuity of normal scientific research that most required explanation. Kuhn's own solution to this problem was an overly static treatment of normal science, one that minimized internal change (Nickles, forthcoming).

14.5 Power Laws and Their Implications

A simple form of a power law distribution is $N(x) = ce^{-\delta x}$, which means that the number of events N of size x or greater equals a constant c times the exponential function. δ is also a constant, a scaling parameter. Power laws[12] and their distributions are scale invariant or "scale free." As indicated, earthquakes provide a familiar example. According to the Gutenberg-Richter law in its simplest form, $N(m) = ce^{-m}$, where m is the magnitude of the earthquake. If we take the logarithm of both sides of power law equations, we get the formula of a straight line as shown in Fig. 14.2. Thus a straight line on a log-log graph is the signature of a power law.

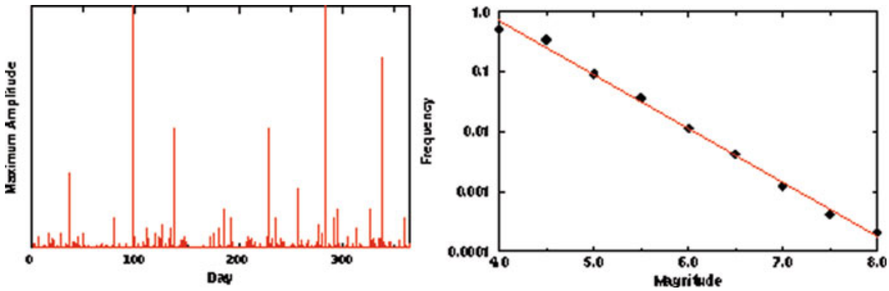

Fig. 14.2 On the *left* is the picture of a time series of earthquakes, plotted according to size. On the *right* is the corresponding log-log graph. Note that the slope parameter δ in this case is approximately –1. Apparently, the slope of the graph varies a bit, according to geographical region, but it is always around –1. From http://simscience.org/crackling/Advanced/Earthquakes/GutenbergRichter.html

[12] A generic form of common power laws is $f(x) = cx^{\delta} + o(x^{\delta})$, where c is the constant, δ is the scaling factor, as before, and o is an asymptotically small function that captures small deviations or uncertainties.

Scientists and other observers long thought of many natural phenomena as the product of large numbers of small, independent events, uncorrelated as to "direction." More recently, this view has begun to give way to a conception of nature as a more highly integrated, complex system than even the "mechanistic universe" of modern physics implies. Especially since the 1960s, power laws have been turning up seemingly everywhere—at least everywhere that complex systems are found—with biological organisms, human artifacts, and social systems being the primary examples. For example, the metabolic rate of biological organisms, from smallest to largest, possibly fits a power law (but see Downs et al. 2008). Ditto for heart rates and so on and on. Hungarian physicist Albert-László Barabási and his group have discovered many (alleged) power laws governing the Internet and various human social networks such as scientists working in a particular specialty area. Power laws turn out to characterize physico-chemical processes at phase transition points (critical points), as when water is about to freeze or turn from liquid to steam or when a piece of iron is about to become magnetized.

The frequent appearance of power laws in complex systems and the importance of phase transitions in many fields raises the interesting question of whether there is a common mechanism behind them, a deep structure underlying the natural phenomena about which we can formulate a general theory. Barabási is convinced that there is such a science of order and connectivity, a topic to which I return in Section 14.8.

14.6 HOT Versus SOC

Several scientists and mathematicians have attempted to formulate a general theoretical approach to complex systems. Given that complexity of one sort or another appears across many scientific disciplines, especially in the biological and social sciences and engineering, a trans-disciplinary theory of complex systems would be an enormous achievement. However, there is wide disagreement about the prospects for such a theory and even about what a complex system is (Gershenson 2008). In this section I briefly compare the approaches of Per Bak's self-organized criticality (SOC) with Carlson and Doyle's highly optimized tolerance (HOT). In the following sections I look more closely at the work of Barabási.

From the 1980s the late Danish physicist Per Bak argued for the importance of "self-organized criticality" (SOC) as the successor to early work on autocatalytic systems and other forms of self-organization at the edge of chaos (EOC) proposed by Stuart Kauffman (1993) and others. Although Bak worked with Kauffman for a time at the Santa Fe Institute, he denied that Kauffman's well-known NK model attains true criticality.[13] Bak's central claim, defended at length in *How Nature*

[13] See Kauffman (1993), Chapter 2 *et passim*. The NK model is Kauffman's start toward "a statistical mechanics of fitness landscapes" (1993, p. 40). "N refers to the number of parts of a system—genes in a genotype, amino acids in a protein, or otherwise. Each part makes a fitness contribution which depends upon that part and upon K other parts among the N."

Works (1996), is that all truly complex systems exist in a critical transition state between order and chaos. Ordinary chaos theory cannot explain complexity, he said. All and only complex systems exhibit his self-organized criticality. His favorite model was the sand pile. Grains of sand are dropped one by one, forming a pile. Eventually, the slopes become steep enough that one additional grain will start an avalanche, usually a small one but occasionally a very large one. When a histogram of avalanches from smallest (displacement of a single grain of sand) to the largest (when perhaps a third of the pile slips away) is plotted, the result is a straight-line, log-log graph, signifying the working of a power law.

Power laws are pervasive in biological nature because evolutionary adaptation drives systems to a critical point, Bak argued. "Biological evolution is a self-organized, critical phenomenon" (1996, p. 150). (But, as noted above, some nonbiological, non-intentional systems, including growing sand piles, also exhibit this sort of behavior, namely at critical points where phase transitions occur.) Moreover, for Bak the power laws indicate a scale-free phenomenon that signals self-similarity and that possesses a fractal signature. Bak explicitly rejected models that are specially "tuned" by an outside designer to achieve critical states. That's where the 'self-organized' comes in: the systems must themselves move toward the critical state, where a certain sort of equilibrium obtains.

Carlson and Doyle (among others such as physicist Geoffrey West, former president of the Santa Fe Institute) contend that Bak's account of complexity is now dated. Carlson and Doyle naturally claim that their HOT supersedes Bak's SOC. They insist that massive empirical information from the physical, biological, and engineering-control worlds shows that SOC theory does *not* capture "how nature works." They disagree with Bak even about what counts as a complex system. Bak clearly wanted to include not only biological but also purely physical systems, systems that manifestly involve no design of any kind. But in throwing out design as a necessary condition, contend Carlson and Doyle, Bak threw out the baby with the bathwater. Their HOT systems, after all, are highly designed, either by biological evolution or by human engineering.

> Highly Optimized Tolerance (HOT), which links complexity to robustness in designed systems, arises naturally through Darwinian mechanisms. ... [R]obustness tradeoffs [are] a mechanism that drives complexity in biology. (Zhou et al. 2002, p. 2049)

> [The model is] sufficiently general that it could be equally well motivated by competition and evolution in other settings, such as between technologies and companies in an economic setting. (Ibid., p. 2050)

While both SOC and HOT differ from statistical mechanical accounts of power laws by linking the power laws to internal structure, Carlson and Doyle (1999) go on to locate HOT at the opposite extreme from SOC, in several ways. (a) A sand pile is boringly self-similar, a mere aggregate, but a HOT system is not. The specialized, highly-engineered systems of a Boeing 777 resemble neither each other nor the airplane as a whole. Ditto the specialized subsystems of biological organisms.

(b) Robust, optimized natural and artificial systems designed by a selection process or by engineers do not necessarily exist at a point of criticality between order and chaos. In general, HOT states are not critical states. (c) HOT systems typically optimize several parameters at once, not just one as in Bak's models. (d) Thus robustness is an *emergent* property of the growth of complex systems, whether these systems are designed by human engineers or by natural selection. (e) The power laws of HOT are "steeper" than those of SOC.

> [T]he HOT power laws are steeper and extend to larger event sizes than the critical power laws, which are very flat. Large events at criticality are fractal, resulting in no macroscopic losses in the limit of large lattices. This is in contrast to both our model and the fossil record, which show losses that are a large fraction of the total organisms or species. (Zhou et al. 2002, p. 2054)

Carlson and Doyle note the policy implications of our choice of complexity model:

> There is much at stake in this debate. If ecosystems are in a SOC/EOC state [i.e., state of self-organized criticality/edge-of-chaos—TN], then observations of massive species extinctions and global warming could be attributed to the natural behavior of the system. In this scenario, large fluctuations emerge and recede as a natural consequence of the internal dynamics, and would not be attributed to man made causes. This would support a policy in which humans could be relatively cavalier about their interactions with the environment, because the system would be fluctuating as observed regardless of our behavior. Alternatively, if ecosystems are in a HOT state then we expect the system to be robust, yet fragile. Heavy tailed distributions are expected, but the system is also hypersensitive to new perturbations that were not part of the evolutionary history. (Carlson and Doyle 1999, p. 1426)

Barabási likewise rejects Bak, saying:

> Networks are not en route from a random to an ordered state. Neither are they at the edge of randomness and chaos. Rather, the scale-free topology is evidence of organizing principles acting at each stage of the network formation process. There is little mystery here, since growth and preferential attachment can explain the basic features of the networks seen in nature. No matter how large and complex a network becomes, as long as preferential attachment and growth are present it will maintain its hub-dominated scale-free topology. (2002, p. 91)

We'll return to preferential attachment in a moment. My present observation is that while Bak's sand pile model was a dynamic approach to self-organized complexity, apparently it was not dynamic enough. According to its critics, it misses crucial details of growth. It pays insufficient attention to both the ongoing processes of generation and corruption and the specialization of the resulting structure. To be fair, Bak claims (p. 150) that biological evolution moves organisms toward what he terms a critical state, but, as we have seen, his critics deny that sandpile type models can really capture this.

14.7 A Bit of Network Theory

As mentioned above, Barabási claims that his research group has founded a new, rigorous scientific inter-discipline of complex systems, one that unveils the deep structure common to all such systems, from biology to business.[14] The brilliant Hungarian mathematicians Paul Erdős and Alfréd Rényi gave the field a start in the 1950s with a series of papers on the connectivity characteristics of random graphs, that is, graphs whose nodes or vertices are connected randomly. Adding links to such graphs, one by one, between randomly chosen nodes, brings the net to a critical state in which adding a couple more links suddenly transforms it into one that is fully connected, i.e., each node can now reach every other node through a path consisting of a series of links or "edges." (One thinks here of the "tipping point" phenomenon popularized by Malcolm Gladwell, 2000.) Conversely, subtracting just a few links randomly can turn a well connected network into a highly fragmented one. This rapid increase or decay is exponential. Thus the normal curve (more precisely, the Poisson distribution in this case) with its exponential drop-off is the signature for random networks just as for random, independent events in nature.

Later Steven Strogatz at Cornell and his former student, Duncan Watts of Columbia University, became interested in applying graph theory to real-world systems. They were especially interested in "small world phenomena" in which almost any two people or points are at most a few links apart but in which there are local clusters of highly interconnected nodes. Efficiency in many systems requires a small number of "degrees of separation," but, for that very reason, such systems are also vulnerable to epidemic contagion. To achieve small worlds with clustering, Strogatz and Watts started from a ring lattice of nodes connected only to their neighbors as in Fig. 14.3, then randomly added a few long-distance links across the circle (Strogatz 2003; Watts 1999). The remarkable result was that just a few additional links tremendously reduced the average degree of separation of the network. By combining the order of the initial lattice with a dash of randomness, they achieved small world networks that were also clustered. However, the element of randomness was enough for them to retain an exponential signature.

The third stage in this development was accomplished by Barabási and his students (especially Réka Albert and Hawoong Jeong), who concluded that real networks are usually even more clustered, containing a few "hubs" with a very large number of connections. This network topology proved to be far more robust than the others. Since the vast majority of nodes are not hubs, now randomly disconnecting a majority of links would produce nothing worse than a graceful degradation of the network. And the signature of these network topologies is . . . power laws and their fat-tailed distributions! In other words, these networks are scale free (up to the limit

[14] There is already a large literature, both technical and popular, on these developments. Newman et al. (2006) is a collection of many of the most influential papers to that date. See also the Carlson and Doyle articles, Watts (1999), and Jen (2005). Among the popular or semi-popular works, see Buchanan (2002, 2007), Miller and Page (2007), and Barabási (2002). The websites of many of these people often contain additional resources.

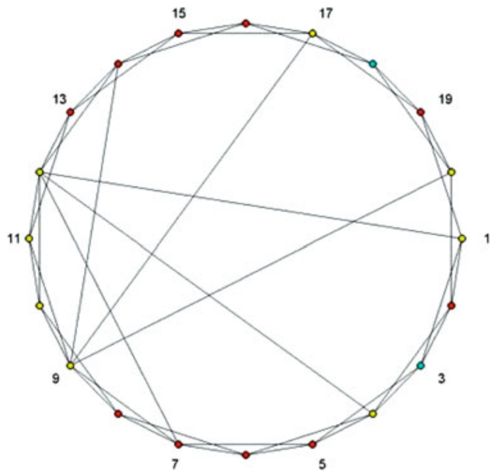

Fig. 14.3 A graph of the Strogatz–Watts type that combines the features of local clustering (by starting from a periodic ring lattice in which only nearest neighbors are connected) and randomness (by adding a few random connections)[15]

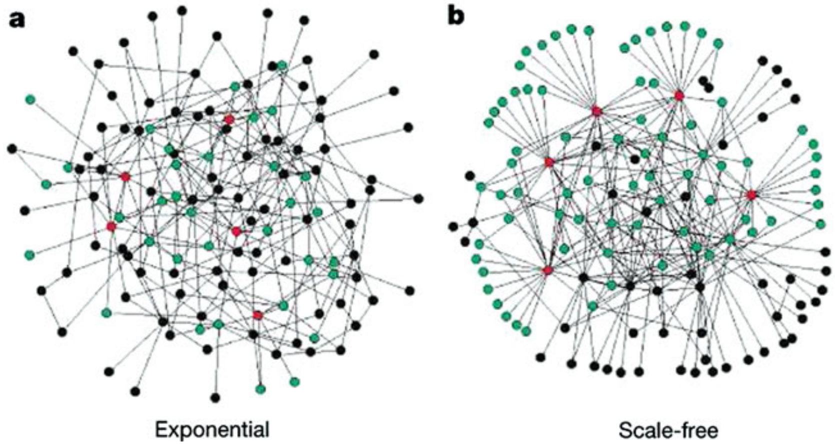

Fig. 14.4 (**a**) The exponential [random] network is homogeneous: most nodes have approximately the same number of links. (**b**) The scale-free network [with hubs] is inhomogeneous: the majority of the nodes have one or two links but a few nodes have a large number of links, guaranteeing that the system is fully connected. *Red*, the five nodes with the highest number of links; *green*, their first neighbours. Although in the exponential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are reached, demonstrating the importance of the connected nodes in the scale-free network. Both networks contain 130 nodes and 215 links. The lighter nodes in greyscale are the green ones. The red and black ones are not distinguishable. [The network diagrams are taken from Albert et al. 2000, p. 379.][16]

[15] The graph can be found at www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=4206&objectType=file.

[16] For the color version, the reader may consult Albert et al. (2000) or http://www.computerworld.com/s/article/75539/Scale_Free_Networks.

of their size). They look the same at all scales. Figure 14.4 is a comparison diagram from Barabási's group (Albert et al. 2000), with their own explanation.

The Barabási group and other experts around the world soon found numerous examples of this sort of network in the biological, technological, and social worlds. According to them, biological food webs, gene networks, and protein networks have this structure. So do commercial airline routes, national electrical grids, and the Internet. And in social worlds, AIDS spreads via such a network, which is extremely efficient in connectivity yet highly robust to random failure.

Is there still a fragility compromise? Yes there is. Where is such a network vulnerable to attack? Answer: in the hubs. While scale-free networks are robust to random failure and also to various kinds of upgrading, as, for example, with the Internet (Willinger and Doyle 2005), they can fail disastrously under targeted attack. Attacks on the hubs can quickly produce a fragmented residue of a once well-connected network. That is how a few "crackers" (the label given to evil hackers as opposed to good hackers) have been able to bring the Internet to a halt with some simple viruses and worms. The net structure allows bad stuff to spread just as efficiently as good stuff.

The happy side of this worrisome feature is that sometimes we do want to destroy networks. For example, Barabási and his colleagues have suggested that the most effective way to treat AIDS and similar diseases, given the expense of the available drugs, is to go after the hubs. The hubs are Malcolm Gladwell's "connectors" (Gladwell 2000, Chapter 2), the people who have done far more than any others to spread the disease. This strategy might also apply to attacking international terrorism.

The fat tails phenomenon suggests another bit of good news for our efforts to understand innovation more endogenously (see Section 14.10 below). Just as systemic failure can always take us by surprise, so can systemic success, including unimagined successes of large magnitude. The trouble is, there are a lot more ways to fail than to succeed. And, again, we should not confuse the two types of failure—getting efficient results that we do not want (such as the spread of a virus) versus disintegration of the network itself.

A second word of caution is in order here. Although these investigators have made great progress in exploring the mathematical and causal structures that underlie networks, the nets themselves are much too simple to capture a lot of rich interaction. In real-world networks, as opposed to their human representation in idealized models, not only are the nodes typically directed (corresponding to a directed graph) but also there are nodes and links of many different kinds. For example, not all human-to-human links are equal. This becomes obvious immediately when we play the "degrees of separation" game. "How many degrees of separation do you have from Einstein?" Well, that obviously depends on what counts as a direct *connection* with (i.e., one degree of separation from) another person. I once shared an elevator ride with Kurt Gödel, who was one degree from Einstein; but as a young graduate student I was too shy to say a single word. Does that count as a connection? Am I only one degree of separation from Gödel and two from Einstein? Obviously not if that means any sort of personal acquaintance, let alone being on a "first name basis," which is the criterion sometimes used. Newman (2001) defines two scientists as directly connected only if they have co-authored a paper.

Nonetheless, the simple networks explored so far have shed considerable light on the topic of robustness and fragility and tend to confirm Perrow's original insight that we must always consider tradeoffs. This approach has given real direction to research programs across the natural and human sciences, engineering, and business.

However, a third important caution is that critics such as John Horgan and Evelyn Fox Keller argue that the enthusiasm for a general complexity theory is premature and that claims for the ubiquity of power laws and scale-free networks are unjustified, amounting to something of an intellectual fad (Horgan 1995; Keller 2005; Downs et al. 2008; Mitchell 2009). This is the primary fear of complexity theorists themselves (Gershenson 2008). Sornette (2003, p. 208) points out that power law distributions are difficult to extract from data sets, given their similarity to other distributions, that many different mechanisms can produce power laws, and that log-periodic features are sometimes more reliable indicators of the underlying mechanisms. There are, of course, many distributions with exponential terms, including one with the technical name "the exponential distribution." For simplicity, I focus on power law versus normal distributions.

14.8 Dynamic Networks

Normal distributions signal randomness, a lack of structure, in some cases the product of entropic processes that break down structure. Conversely, then, we are invited to look at processes that *produce* structure for the origin of power law distributions. But not all structures give rise to power laws. Simply being non-Gaussian is not enough. Why the seemingly pervasive existence of scale-free power-law distributions? And does the explanation have something to do with robustness and/or complexity? Carlson, Doyle, West, and Barabási all think so.

For Bak, as we have seen, the explanation is nothing more special than what his sand pile model suggests. Simple aggregation of ordinary events can produce critical states with nonlinear consequences. For West, who is closer to the Carlson-Doyle camp, power laws themselves imply the existence of a robust design or mechanism that produces them, subject to physical constraints.

In effect, Carlson and Doyle develop Perrow's old theme that the drive for robustness generates spiraling complexity vulnerable to unexpected, cascading failure as signaled by power law distributions with their fat tails. Carlson and Doyle add a second drive—for optimal throughput. And the mechanism in both cases is natural selection (or a human design analogue), which always involves a compromise among many factors.

Meanwhile, Barabási proposes a more detailed *growth* model for networks exhibiting power laws. The two principles in play here are "the early bird principle" and "the rich-get-richer principle," resulting in a network with distinct hubs. In the growth of such a network, being an early node increases the probability of becoming a hub simply because later links must be made to existing nodes. But preferential attachment also plays a role in the development of many systems, meaning that linking with an already well-linked node is more probable than linking with a relatively isolated node. The rich get richer.[17]

A recent move in this particular debate involves a disagreement between the HOT model and the preferential attachment model. The HOT team claim that preferential attachment does not fit the empirical data as well as their model does. To mention one example: D'Souza et al. (2007) state that preferential attachment stands in the Pareto, Polya, Zipf, and Simon tradition, whereas the optimization approach goes back to Mandelbrot's statistical study of language.[18] D'Souza et al. propose to reconcile the two approaches with their "tempered preferential attachment model," a model that attempts to explain preferential attachment in terms of a cost function rather than simply assuming it. In their model, nodes can become saturated with links, leading to an exponential cutoff, and not all attempts to create new nodes are successful. The overall result is that a tempered preferential attachment emerges from the net formation process subject to the cost constraints.

That is only one example. Various investigators have recently pointed out that there are many mechanisms for producing power law distributions, that preferential attachment is only one of them (and is not really original with the Barabási group). Accordingly, anyone who claims the evidence for a power law distribution proves an underlying preferential attachment mechanism is guilty of the fallacy of affirming the consequent. (I do not say that the Barabási group make such a strong claim.) So, again, we must proceed cautiously here and realize that gestures toward real-world applications are highly conjectural.

14.9 Application to a Kuhnian Model of Science

To apply these ideas in detail to models of scientific research would require a book examining citation networks, actor networks of Bruno's Latour's sort (Latour 1987), semantic networks, and so on. This is obviously not the place to tackle the major task of working the growing literature into an explicit model. Instead I shall point out possible connections to a broadly Kuhnian conception of mature science. Kuhn's model is problematic, of course, but so are all such models. For present purposes I shall stick close to Kuhn's, as the one best known. My analysis suggests that future study of various types of scientific networks and their transformation will offer support for some aspects of Kuhn's model while undermining others.

In *The Structure of Scientific Revolutions* (1962), Kuhn already offered a proto-complexity model that fits my extension of Perrow's thesis to scientific inquiry. Research scientists seek to design structures (research programs, theories, models,

[17] Both principles would seem to play a role in Wimsatt's conception of generative entrenchment as applied to developing systems, either biological or humanly devised.

[18] Among other attempts to explain some power laws, see Fabrikant et al. (2002). The study of phase transitions at critical points is another locus of such efforts.

experimental systems, etc.) that are robust to anticipated kinds of failure. The conviction that a more robust account of their scientific domain is not only possible but also now accessible to inquiry is what drives Kuhnian normal scientists to design increasingly intricate and esoteric structures and practices, as the other chapters in this volume attest. What Popper considered the very core of scientific research—formulating bold conjectures followed by vigorous attempts to falsify them—Kuhn recast as threats to be controlled. A well-designed Kuhnian paradigm is robust to both threats. Normal scientists who are too bold will be disciplined by the community, Kuhn said; and Popper's "falsifications" of major principles are really only routine anomalies that provide new research puzzles. The basic principles are not up for test in the first place. Kuhnian normal science is far more tolerant of what Popper considered mistakes than is even Popper's account of science.[19] The resulting solidarity of the research community over what counts as legitimate problems and solutions permits not only routine problem-solving success but also productive exploration of new, increasingly esoteric research puzzles.

However, Kuhn then surprised the philosophical community by emphasizing the fragility of paradigms.[20] The very robustness and productivity of a successful Kuhnian paradigm in the indicated respects makes the enterprise increasingly vulnerable to major failure, he said.

> Anomaly appears only against the background provided by the paradigm. The more precise and far-reaching that paradigm is, the more sensitive an indicator it provides of anomaly and hence of an occasion for paradigm change. . . . By ensuring that the paradigm will not be too easily surrendered, resistance guarantees that scientists will not be lightly distracted and that the anomalies that lead to paradigm change will penetrate existing knowledge to the core. The very fact that a significant scientific novelty so often emerges simultaneously from several laboratories is an index both to the strongly traditional nature of normal science and to the completeness with which that traditional pursuit prepares the way for its own change. (1970, p. 65)

The very robustness of ordinary science (whether or not precisely Kuhnian) makes it increasingly vulnerable to revolutionary overthrow. What is it about a mature science that makes it ripe for revolution? A cognitive psychological point is that strong focus by the community on narrow, esoteric matters provides guidance to the community at the frontier of research but blinds it from other things. A social

---

[19] Popper attempted to remove the fear of making mistakes. His motto was "We learn from our mistakes." In his methodology of scientific research programs, Lakatos (1970) followed Kuhn in the respect mentioned in the main text while retaining some Popperian elements. Kuhn (1970) objected to Popper's talk of falsifications as mistakes.

[20] In the chapters on normal science, Kuhn invites us to look at scientific work from the point of view of the normal-scientific practitioner, who, according to Kuhn, is convinced that s/he is uncovering the truth about the world. In the chapters on scientific revolutions, Kuhn invites us instead to look at the history of science from above and to note the contingency involved in the revolutionary passage to a new paradigm. As a rough generalization, philosophers have tended to take a normal scientific, realist view whereas sociologists have distanced themselves from the normal science perspective. In my opinion, contrary to Kuhn, taking the normal scientists' viewpoint does not require conviction that the paradigm is on the track to a final truth about the world.

psychological point is that routine anomalies become more serious over time as they resist the attempts of the best people to resolve them. Confidence in the resources of the paradigm falls rapidly (unless, I would add, there is compensating, rapid progress on other fronts). The emergence of a possible alternative approach can also produce a crisis by amplifying the importance of extant anomalies (Kuhn 1970, p. 86). But the Kuhn quotation suggests an additional, non-psychological mechanism. As a science matures, the linkage between its components becomes more complete and more entrenched and the research more systematic. Thus a minor measurement discrepancy or a new discovery that does not quite fit can now penetrate deeply. The maturation of the paradigm has given the anomaly more leverage. It is now more threatening. The possible error that it represents can no longer be contained. It increasingly propagates through the system, in ways somewhat reminiscent of Descartes' system, discussed above. Previously, it could be bracketed, perhaps treated by analogy with the "God of the gaps" tactic familiar from theology; but now the gaps have closed.

Per Bak would say that normal science has reached a critical state in which even a "normal" result could trigger a revolution, and that this becomes explicit in a Kuhnian crisis.[21] However, although normal science is cumulative, according to Kuhn, there is far more structure to the accumulation than Bak's sandpile model allows. For many analysts a sandpile epitomizes a mere aggregation rather than a complex system (Wimsatt 2006). For this and other reasons, the HOT model of Carlson and Doyle fits Kuhnian science better. For the goal of science is to maximize problem-solving productivity while eliminating normal (expected) kinds of error—and to do so precisely by improving the design of theory and data structures. Moreover, the development of normal science would seem to be broadly evolutionary (Nickles, forthcoming).

We should credit Bak with two other important insights that Carlson and Doyle also appropriate and incorporate in the HOT model. One is the point already stated above: that in a highly mature, rigorous, science even a seemingly ordinary result can trigger a cascade of developments that lead to revolution. Scientific work is highly nonlinear in this respect. It does not take a big, revolutionary cause to trigger a process that produces a revolutionary effect.[22] Again, most anomalies begin as

[21] Bak does not mention Kuhnian revolutions, though he does relate his work to other models of transformative change such as Gould-Eldredge punctuated equilibrium and mass extinction (Bak, Chapter 1). Sornette (2003, Chapter 3) contends that the causes of stock market crashes are not ordinary events, that crashes are outliers with a special statistics of their own that call for special explanations. However, his model is not totally different from those under discussion. The underlying processes involve increasingly correlated phenomena of complex systems, driven by positive feedback, that send the system to a critical point, where it becomes unstable. At this point a normal change can tip the system one way or the other. As an emergent phenomenon of a complex system, such a disruption, like a Kuhnian revolution, is holistic. It cannot be analyzed into component parts.

[22] This point is nearly explicit, however, in Kuhn's other major work, his history of the quantum theory. (See my discussion of his Planck case in Nickles, 2009 and forthcoming.) Critics complained that Kuhn failed to integrate his quantum history with the model of *Structure* (Klein et al. 1979).

small disruptions that normal scientists have every reason to believe they can handle with the resources available to them. Roughly speaking, the more mature (robust) the science, the greater the nonlinearity, since small discrepancies now have more leverage.

Bak's other insight is that these disruptions occur on all scales. Applied to Kuhn, this suggests that there is no principled distinction between normal and revolutionary science, that normal science is more dynamic than Kuhn's account allows, and that revolutions are simply normal disturbances writ large (cf. McMullin 1993, Wray 2007).[23] But perhaps a better way to express the scaling point is this. Kuhn states that there are small revolutions within specialty and subspecialty fields as well as large, highly visible revolutions. A small revolution may look like cumulative change to those practitioners working in other fields who notice it at all. If Kuhn is correct, this suggests that the structure of the general field is modular with the connections between some of the specialty areas and even between them and the core rather weak. In such a case the large field is, in Herbert Simon's term, nearly-decomposable, at least to some extent (Simon 1981; Wimsatt 2007, Chapter 9).

We can also notice a connection to Barabási's preferential attachment model. In Kuhn's account of normal science (especially as amplified in his "Postscript—1969"), exemplars come to function as hubs. New research puzzles are solved by relating them to one or more genealogies of puzzles and solutions that the community takes as exemplary. Accordingly, the discovery that an entrenched exemplar is defective or that it has exhausted its ability to yield new problem-solving insights sends shocks through much of the corresponding normal science. Further elaboration of this point would bring in Wimsatt's aforementioned work on generative entrenchment. Given the historical contingency of which specific exemplars are, in effect, selected as the "early birds," the ensuing scientific work under that paradigm is likely to remain contingent in important respects owing to historical path-dependence, even though scientific practitioners do often succeed in reworking older material in such a way as to eliminate some contingencies (Nickles 1997). This point directly connects with Soler's concern with the historical contingency versus inevitability thesis in her chapter in this volume.[24]

---

[23] It is then open to a Kuhnian to reply that genuine revolutions are precisely those disturbances that could *not* be contained within the bounds of normal science and that resulted in overturning the old approach, that making the distinction a matter of degree violates the hierarchical nature of his model. One response would be that even Kuhn, qua historian, agrees that classical mechanics went through several phase changes between Newton and Einstein, changes that transformed it almost beyond recognition as it incorporated the so-called Baconian sciences; later adopted the latest Lagrangian, Hamiltonian, and other mathematical techniques; rejected the ideas of action-at-a-distance and that all forces are central forces; became statistical-probabilistic, etc. The very concept of mechanics was transformed in the process. So why count all of this as normal science?

[24] See Soler (Chapter 10, Section 10.16) as well as Soler (2000, 2004). Kuhn (1962) stated that paradigm change is almost inevitable given the unavoidable contingency of its formation. After all, at the frontiers of a new domain of research it is most obvious that scientists cannot yet know much about the structure of that domain. It is thus exceedingly improbable that a given paradigm will

Are mature branches of science that lack the theoretical integration of mechanics equally subject to Kuhn-style revolutions? The chemical and biological sciences, for example, possess a less theory-centered structure. Are they therefore less subject to catastrophic failure of the relativity or quantum theory variety and thus supportive of a more robust scientific realism in their domains? This question is worth further exploration. On the other hand, inspired by Kuhn's own talk of exemplars as practically making a theory structure unnecessary, writers such as Giere (1988, Chapter 3, 2008), Teller (2001, 2008) and Rouse (2003) adopt a more exemplar-centered than theory-centered conception of Kuhnian scientific practice, even in mechanics. On this view a "theory" is really a collection of models.

Finally, the network approach helps us to make sense of how Kuhn can pass so quickly from talk of incommensurability (e.g., of relativity theory with classical mechanics) to the following claim:

> We may even come to see [the relativity revolution] as a prototype for revolutionary reorientations in the science. Just because it did not involve the introduction of additional objects or concepts, the transition from Newtonian to Einsteinian mechanics illustrates with particular clarity the scientific revolution as a displacement of the conceptual network through which scientists view the world. (*Structure*, p. 102)

Talk of the same concepts is odd, since Kuhn has just made his strong claim that meaning change prevents literal limit relationships between the relativity theory and classical mechanics. But we can understand what he is getting at in terms of networks. In the background is a quasi structuralist account of meaning according to which the individual linguistic units are meaningless in themselves (or have arbitrary meaning), their technical meaning deriving from their place in the structure or network, combined with the logical empiricist idea of implicit definition. With the coming of relativity theory, the linkages among the concepts changes, thereby producing a holistic change in the meaning of the concepts themselves, a change that cannot be analyzed as a piecemeal change in a single "definition" or two. (There will be a corresponding cognitive change for those who understand the language in the new way.) Kuhn is treating the concepts as syntactic nodes the meaning of which depends on their place in the network.[25] And it is this claim and perhaps also the behavioral economics of the Kuhnian scientific community that suggest that self-organizing systems may be in play. (The earlier point about the nonlinear vulnerability of mature normal science is rather different from the

be able to anticipate future results so as to get everything right. Thus Soler's treatment of contingency nicely complements my own about maturation as increasing vulnerability to transformative change in the system. Sociologists of science from Latour and Woolgar (1979), Knorr-Cetina (1981), and Pickering (1984) to the present have given far more attention to the contingency of scientific decisions than have philosophers of science.

[25] This last is a familiar point often made about Kuhn and Feyerabend. Papineau (1979) provides an excellent discussion. Kuhn retained a more limited sort of meaning holism later in his career (Kuhn 2000). Soler (2000, 2004) interprets Kuhn's account of meaning and meaning change in terms of the structuralism inspired by Ferdinand de Saussure's work.

point about self-organization, but ultimately related.) A Kuhnian scientific revolution, like a transformative change of a complex system, is an emergent, holistic macro-phenomenon that cannot be analyzed adequately at the level of its component parts. It remains unclear (at least to me) to what extent a Kuhnian scientific community, with its imitation or herding tendencies, can be regarded as a *self-organizing* complex system in which positive feedback can take the system through a critical point and into a new regime.

14.10 Prospective Robustness

Robustness is a concept typically applied to specific research results, based on their track record of empirical and theoretical support, especially in Wimsatt's (1981) sense that robust results can be derived and/or checked in multiple ways. I have extended the concept of robustness to entire epistemic systems, regarding them as designed problem-finding and problem-solving systems that, when successful in surviving shocks, evolve toward states of increasing robustness in some respects but also increasing vulnerability to failure in other respects. And I have relativized robustness to specific dimensions of failure. Typically, these dimensions of potential failure are somehow "anticipated." Admittedly, this is a difficult idea to apply to biological systems without lookahead, but it is crucially important to human inquiring systems. Accordingly, I want to suggest a further extension of the concept of robustness, to more explicitly include designing for the future, as the network theory we have canvassed suggests. My thought, inspired by Kuhn's, Pickering's, and Wimsatt's work on heuristic fertility, is that a research program (for example) is more robust than another insofar as its long-term prospects for fertile development are better.[26]

It is this prospective "heuristic appraisal" (as I call it) that enables Kuhnian paradigms and Lakatosian research programs to be robust to the anomalies that Popper regards as falsifications. In making science safe for failure with his emphasis on learning from our mistakes, Popper took an important step forward. But it was simultaneously a step backward: in excluding all sorts of ad hoc hypotheses, and in saying that scientists should reject a "falsified" theory, Popper amplified the destructive effects of anomaly. He made scientific work more robust and less robust at the same time and in the same respect, to the failures that he called falsifications.

Unlike nonhuman biological evolution, human inquiry can take advantage of some degree of lookahead, and this prospective orientation marks a major difference between Kuhn and the logical empiricists and Popperians.[27] Traditional

[26] Hume's problem of induction implies that we should be fallibilists about the future, but the problem is so general that it undercuts all enterprises more or less equally. More specific considerations involved in what I call heuristic appraisal provide differential appraisals of future prospects and thereby make a difference (Nickles 2006).

[27] Lakatos (1970) and his students made heuristics an important part of search programs.

confirmation theory and Popper's theory of corroboration treat the epistemic merit of a theory or research program as a function of its empirical track record to date and regard empirical failure as epistemic death (falsification). Kuhn fundamentally disagreed. For Kuhn (1962, 1970) it is prospective fertility that determines the viability of a research program, not past success or failure. Since all decisions are about the future, future prospects are obviously critical. The point is both sociological and technical-scientific. Scientists are attracted to a paradigm in the first place because they can see how to use it to investigate interesting problems and thus contribute to that specialty and, in the process, to build their careers and maintain their self-identity as productive scientists in a specific field.

An indication of Kuhn's departure from the received view is his rejection of the claim that "context of discovery" (as a general label for innovative activities) is not epistemologically interesting. To be clear: Kuhn did not believe there is a logic of discovery or a rigid scientific method of any kind, including a method of justification. However, in his view a good research program provides strong indications of where and how future work is likely to bear fruit. In the best cases, the paradigm practically guarantees that any research puzzle that can be formulated in its terms is solvable by means of its resources, which also explains the rapid drop-off of confidence when failure is evidence.

This prospective versus purely retrospective conception of robustness carries over to smaller-scale units such as an investigator's experimental system (Rheinberger 1999) and even to research proposals. A good system or proposal is one that has a clear direction but also one that is flexible enough to be opportunistically adaptable in various ways in response to anticipated possibilities of failure and success. It is not fragile in the sense of falling flat at the first sign of failure. It is one with strong heuristic promise.[28]

14.11 Concluding Summary

I have argued for a broadened conception of robustness, one coupled to fragility and one that takes into account the future, at least where some degree of lookahead is possible. Robustness is not absolute but relative to types or sources of failure. Increasing robustness in one dimension is typically coupled to increased

[28] Although the two are intricately linked (Rescher 1977), we need to distinguish between the retrospective and the prospective (or heuristic) robustness of a product and the robustness of the process that produced it. A product can be robust even though the process that produced it is now regarded as "played out," sterile, unlikely to produce fruitful new results, thus not robust in the heuristic sense. And a process may be evaluated as robust in the heuristic sense even if has not yet produced much in the way of warranted products. Thus we need to interpret my basic expression $R(s,p,d,c)$ generously to allow that robustness of system s in the dimension(s) of future promise might be of degree d great enough to withstand both endogenous perturbations p (anomalies, adjustments to the research system) and less endogenous ones such as the complaint of underdevelopment, compared to the competition.

vulnerability in others. Thus I reject the cumulative fragility-reduction/risk-reduction thesis, a point that at least somewhat undercuts the appeal of some strong realists to the success of (usually recent) "mature science" as the basis of their claims.[29] Major failures may always result from unexpected exogenous events (a large meteorite hitting the earth, a large cut in project funding, a major breakthrough by a competitor), but they may also result from endogenous developments owing to the nonlinearity inherent in complex systems.

Since the same point holds for major successes, this is a step toward an endogenous account of scientific innovation ("discovery"). After all, we do want our inquiry systems (both processes and products) to change in positive ways. In accordance with the literature on complex systems, I have extended robustness attributions beyond individual empirical claims to entire systems, whether natural or artificial, and I have focused on robustness in the sense of ability to respond adequately to shocks. I have suggested that the philosophy of science literature on robustness needs to go beyond standard confirmation theory to include the prospective evaluation of future promise. Thus I should like to shift the emphasis from the dominant, retrospective, analytic epistemological view approach that is most concerned with the degree of justification of claims already on the table, and of the processes that produced them, to an approach more in tune with scientists' own, future-oriented points of view at the creative frontiers of research. The first approach remains important, of course, but equally important is the second; and, as I have attempted to show, there is an interaction between the two. Our evaluations of future prospects are, of course, fragile in obvious ways; but our evaluations of past results are also fragile—in less obvious ways—precisely *because of* the necessity of anticipating future changes!

Acknowledgements Thanks to Léna Soler for organizing the conference on robustness at Nancy 2, to the members of the Poincaré Archives for their hospitality, to the participants, especially Léna, Bill Wimsatt, an anonymous referee, and Gaye McCollum-Nickles, for helpful comments on either my presentation or a previous draft. I am also generally indebted to Andy Pickering for his attention in his publications to what I call heuristic appraisal and to his pragmatic outlook on the sciences generally. For discussion of Kuhn I am indebted to my students, Jared Ress and Jonathan Kanzelmeyer.

References

Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 2000. "Error and Attack Tolerance of Complex Networks." *Nature* 406(27 July):378–81. Reprinted in Newman et al. (2006), 503–6.

Bak, Per. 1996. *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus, Springer.

Barabási, Albert-László. 2002. *Linked: The New Science of Networks*. Cambridge, MA: Perseus.

Bridgman, P.W. 1927. *The Logic of Modern Physics*. New York: Macmillan.

[29] A detailed discussion of scientific realism is not possible in this already long chapter. For the importance of appeals to the success of mature science in recent defenses of realism, see, e.g., Laudan (1981), Leplin (1984), and Psillos (1999).

Buchanan, Mark. 2002. *Nexus: Small Worlds and the Groundbreaking Theory of Networks*. New York: Norton.

Buchanan, Mark. 2007. *The Social Atom: Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks Like You*. New York: Bloomsbury.

Calvin, William. 2002. *A Brain for All Seasons: Human Evolution and Abrupt Climate Change*. Chicago: University of Chicago Press.

Carlson, Jean, and John Doyle. 1999. "Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems." *Physical Review* E60:1412–27.

Carlson, Jean M., and John Doyle. 2002. "Complexity and Robustness." *Proceedings of the National Academy of Science* 99:2538–45.

Downs, Cynthia, J.P. Hayes, and C.R. Tracy. 2008. "Scaling Metabolic Rate with Body Mass and Inverse Body Temperature: A Test of the Arrhenius Fractal Supply Model." *Functional Ecology* 22:239–44.

Doyle, John, et al. 2005. "Robustness and the Internet: Theoretical Foundations." In Jen (2005), 273–85.

D'Souza, Raissa, C. Borgs, J.T. Chayes, N. Berger, and R. Kleinberg. 2007. "Emergence of Tempered Preferential Attachment From Optimization." *Proceedings of the National Academy of Sciences* 104:6112–7.

Duhem, Pierre. 1954. *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press. Translated from the French edition of 1914.

Fabrikant, Alex, Elias Koutsoupias, and Christos Papadimitriou. 2002. "Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet." In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming (ICALP)*, edited by P. Widmayer, et al., 110–22. Berlin: Springer.

Feynman, Richard. 1965. *The Character of Physical Law*. Cambridge, MA: MIT Press.

Gershenson, Carlos, ed. 2008. *Complexity: 5 Questions*: Automatic Press/VIP/Vince.

Gertstein, Marc. 2008. *Flirting with Disaster: Why Accidents Are Rarely Accidental*. New York: Union Square Press.

Giere, Ronald. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Giere, Ronald. 2008. "Comment on Teller: Incommensurability from a Modeling Perspective." In Soler et al. (2008), 265–9.

Gladwell, Malcolm. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Boston: Little Brown.

Hindo, Brian. 2007. "At 3M, a Struggle Between Creativity and Efficiency." *Business Week*, June 11.

Horgan, John. 1995. "From Complexity to Perplexity." *Scientific American* 272(June):74–9.

Jen, Erica, ed. 2005. *Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies* (Santa Fe Institute Studies in the Sciences of Complexity). Oxford: Oxford University Press.

Kauffman, Stuart. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

Keller, Evelyn Fox. 2005. "Revisiting 'Scale-Free' Networks." *BioEssays* 27:1060–8.

Klein, Martin, Abner Shimony, and Trevor Pinch. 1979. "Paradigm Lost?" *Isis* 70:429–40.

Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge*. Oxford: Pergamon Press.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. 2nd ed. with "Postscript—1969", 1970. Chicago: University of Chicago Press.

Kuhn, Thomas. 1970. "Logic of Discovery or Psychology of Research?" In Lakatos and Musgrave (1970), 1–20.

Kuhn, Thomas. 2000. *The Road Since Structure*, edited by James Conant and John Haugeland. Chicago: University of Chicago Press.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In I. Lakatos and A. Musgrave, 91–195.

Lakatos, Imre, and Alan Musgrave, eds. 1970. *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills, CA: Sage.

Laudan, Larry. 1981. "A Confutation of Convergent Realism." *Philosophy of Science* 48:19–49.

Leplin, Jarrett, ed. 1984. *Scientific Realism*. Berkeley, CA: University of California Press.

Levin, Simon, et al. 1998. "Resilience in Natural and Socioeconomic Systems." *Environment and Development Economics* 3:225–36.

Martin, O.C., and Andreas Wagner. 2008. "Multifunctionality and Robustness Tradeoffs in Model Genetic Circuits." *Biophysical Journal* 94:2927–2937.

McMullin, Ernan. 1993. "Rationality and Paradigm Change in Science." In *World Changes: Thomas Kuhn and the Nature of Science*, edited by Paul Horwich, 55–78. Cambridge, MA: MIT Press.

Miller, John, and Scott Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press.

Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. New York: Oxford University Press.

Newman, Mark. 2001. "Scientific Collaboration Networks." *Physical Review E* 64, paper 016131.

Newman, Mark, and Richard Palmer. 2003. *Modeling Extinction*. Oxford: Oxford University Press.

Newman, Mark, Albert-László Barabási, and Duncan Watts, eds. 2006. *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Nickles, Thomas. 1997. "A Multi-Pass Conception of Scientific Inquiry." In *Danish Yearbook of Philosophy 1997*, vol. 32, edited by Stig Andur Pedersen, 11–43. Copenhagen: Museum Tusculanem Press.

Nickles, Thomas. 2003. "Evolutionary Models of Innovation and the Meno Problem." In *International Handbook on Innovation*, edited by Larisa Shavinina, 54–78. Amsterdam: Elsevier.

Nickles, Thomas. 2006. "Heuristic Appraisal: Context of Discovery or Justification?" In *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*, edited by Jutta Schickore and Friedrich Steinle, 159–82. Dordrecht: Springer (Archimedes Series).

Nickles, Thomas. 2009. "Scientific Revolutions." *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/scientific-revolutions/.

Nickles, Thomas. Forthcoming. "Some Normal Scientific Puzzles." In *Kuhn's "The Structure of Scientific Revolutions" Revisited,* edited by Theodre Arabatzis, and Vasso Kindi. London: Routledge.

Papineau, David. 1979. *Theory and Meaning*. Oxford: Oxford University Press.

Perrow, Charles. 1972. *Complex Organizations: A Critical Essay*. 3rd ed. New York: McGraw-Hill.

Perrow, Charles. 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Perrow, Charles. 2011. *The Next Catastrophe*. Princeton, NJ: Princeton University Press.

Pickering, Andrew. 1980. "Exemplars and Analogies: A Comment on Crane's Study of Kuhnian Paradigms in High Energy Physics" and "Reply to Crane." *Social Studies of Science* 10:497–502 and 507–8.

Pickering, Andrew. 1984. *Constructing Quarks*. Chicago: University of Chicago Press.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Popper, Karl. 1963. *Conjectures and Refutations*. London: Routledge.

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Putnam, Hilary. 1962. "What Theories Are Not." In *Logic, Methodology and Philosophy of Science*, edited by Ernest Nagel, Patrick Suppes, and Alfred Tarski, 215–27. Palo Alto, CA: Stanford University Press. Reprinted in Putnam's *Mathematics Matter and Method: Philosophical Papers*, vol. 1, 2nd ed., Cambridge: Cambridge University Press, 1979.

Quine, W.V. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60:20–43. Reprinted with changes in Quine's *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 1953, 20–46.

Rescher, Nicholas. 1977. *Methodological Pragmatism*. Oxford: Blackwell.

Rheinberger, Hans-Jörg. 1999. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Rouse, Joseph. 2003. "Kuhn's Philosophy of Scientific Practice." In *Thomas Kuhn*, edited by Thomas Nickles, 101–21. Cambridge: Cambridge University Press.

Schumpeter, Joseph. 1942. *Capitalism, Socialism and Democracy*. New York: Harper & Brothers.

Simon, Herbert. 1947 (later editions). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.

Simon, Herbert. 1981. *The Sciences of the Artificial.* 2nd ed. Cambridge: MIT.

Soler, Léna. 2000. "Le concept kuhnien d'incommensurabilité, reconsidéré a la lumière d'une théorie structurale de la signification." *Philosophia Scientae* 4:2(October):189–217.

Soler, Léna. 2004. "The Kuhnian Concept of Incommensurability Reconsidered in the Light of a Saussurian Framework." *Malaysian Journal of Science and Technology Studies* May:51–75.

Soler, Léna, Howard Sankey, and Paul Hoyningen-Huene, eds. 2008. *Rethinking Scientific Change and Theory Comparison: Stabilities, Ruptures, Incommensurabilities?* Dordrecht: Springer.

Sornette, Didier. 2003. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton, NJ: Princeton University Press.

Strogatz, Steven. 2003. *Sync: The Emerging Science of Spontaneous Order*. New York: Hyperion.

Teller, Paul. 2001. "Twilight of the Perfect Model Model." *Erkenntnis* 55:393–415.

Teller, Paul. 2008. "Of Course Idealizations Are Incommensurable." In *Rethinking Scientific Change and Theory Comparison: Stabilities, Ruptures, Incommensurabilities?* edited by L. Soler et al., 247–64. Netherlands: Springer.

Wagner, Andreas. 2005. *Robustness and Evolvability in Living Systems*. Princeton, NJ: Princeton University Press.

Watts, Duncan. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press.

Willinger, Walter, and John Doyle. 2005. "Robustness and the Internet: Design and Evolution." In *Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies,* edited by E. Jen, 231–71. Oxford: Oxford University Press.

Wimsatt, William C. 1981. "Robustness, Reliability, and Overdetermination." In *Scientific Inquiry and the Social Sciences*, edited by M. Brewer and B. Collins, 124–63. San Francisco, CA: Jossey-Bass. Reprinted in Wimsatt (2007), 43–74.

Wimsatt, William C. 2006. "Aggregate, Composed, and Evolved Systems: Reductionistic Heuristics as Means to more Holistic Theories." *Biology and Philosophy* 21:667–702.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Wray, K. Brad. 2007. "Kuhnian Revolutions Revisited." *Synthese* 158:61–73.

Zhou, Tong, Jean M. Carlson, and John Doyle. 2002. "Mutation, Specialization, and Hypersensitivity in Highly Optimized Tolerance." *Proceedings of the National Academy of Science* 99:2049–2054.

Index

L. Soler et al. (eds.), *Characterizing the Robustness of Science*, Boston Studies in the Philosophy of Science 292, DOI 10.1007/978-94-007-2759-5,