FOLK PSYCHOLOGY RE-ASSESSED

FOLK PSYCHOLOGY RE-ASSESSED

Edited by

DANIEL D. HUTTO
University of Hertfordshire, U.K.

and

MATTHEW RATCLIFFE
University of Durham, U.K.

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-5557-7 (HB)
ISBN 978-1-4020-5558-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

TABLE OF CONTENTS

LIST OF CONTRIBUTORS

Kristin Andrews is Assistant Professor of Philosophy at York University, Canada.

Shaun Gallagher is Professor and Chair of Philosophy at the University of Central Florida, USA.

Peter Goldie is Samuel Hall Professor of Philosophy at the University of Manchester, UK.

Peter Hobson is Professor of Developmental Psychopathology at the Institute of Child Health, University College London and the Tavistock Clinic, London, UK.

Daniel D. Hutto is Professor of Philosophical Psychology at the University of Hertfordshire, UK.

Joshua Knobe is Assistant Professor of Philosophy at the University of North Carolina – Chapel Hill, USA.

Martin Kusch is Professor at the Department of History and Philosophy of Science, University of Cambridge, UK.

Victoria McGeer is Lecturer in Philosophy and a member of the Research Faculty, University Center for Human Values, at Princeton University, USA.

Adam Morton is Canada Research Chair in Epistemology and Decision Theory at the University of Alberta, Canada.

Matthew Ratcliffe is Reader in Philosophy at Durham University, UK.

Beata Stawarska is Assistant Professor of Philosophy at the University of Oregon, USA.

Dan Zahavi is Professor of Philosophy and Director of the Center for Subjectivity Research at the University of Copenhagen, Denmark.

1. INTRODUCTION

1.1. FOLK PSYCHOLOGY, THEORY OF MIND AND SIMULATION

The tasks we face in our day to day social lives are quite heterogeneous but many of them make a common demand upon us. They require us to understand and interact with other *people* and, in most social encounters, we exhibit a special sensitivity to our fellow human beings that is quite different from the way we respond to inanimate objects and most other species of organism. Social life is dependent, to a considerable degree, on our ability to understand what is distinctive about human behaviour and to successfully apply that understanding in all manner of situations.

What is central to our ability to interpret one another? A great deal of work in philosophy of mind, cognitive science, anthropology, developmental psychology and a host of other disciplines assumes that, at root, interpersonal interpretation is accomplished through the employment of a 'commonsense' or 'folk' psychology, meaning an 'everyday', rather than 'scientific', appreciation of mindedness. Although there is considerable debate over which cognitive processes support our folk psychological abilities and how those abilities develop during childhood, there is a remarkable degree of consensus concerning what folk psychology consists of. Almost all discussions of the topic begin by stating or presupposing that it is the ability to attribute intentional states, principally beliefs and desires, to other people and perhaps also to oneself, in order to predict and explain behaviour. Davies and Stone's (1995a p. 2) assessment is typical: we "deploy psychological concepts such as belief and desire in predictions and explanations of the actions and mental states of other members of the species". Beliefs, desires and most of the other mental states assigned are taken to be 'propositional attitudes'. In other words, they have the form 'X believes that p' and 'X desires that q', where p and q are propositions that can have any intelligible content you like, such as 'it is raining', 'Paris is the capital of France', 'the cat is under the table' and so forth. It is generally agreed that the 'folk' are realists about propositional attitudes, taking them to be internal states that knit together in complex ways so as to cause actions, although there is much dispute concerning what, precisely, such 'intentional realism' amounts to and whether it is defensible.[1] Propositional attitudes are said to come in two principal types; 'beliefs', which carry information about the world and thus guide action, and desires, which are motivational states that specify goals for action.

Only confusion abounds if we fail to define terms. The most neutral, anodyne definition of 'folk psychology' equates it to the way – whatever it turns out to be – that social beings manage to conduct interpersonal relations (Hornsby 1997, pp. 4–5). As a working definition, this is far too encompassing; nothing useful

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 1–22.

could answer to such a description. At the other end of the spectrum, perhaps the tightest definition of folk psychology would be the practice of making sense of actions in terms of the core propositional attitudes, beliefs and desires, alone. For many, this excludes too much even from our ordinary practice of making sense of one another in terms of reasons. However, those interested in the topic have typically erred on the side of caution, sticking to the tighter definition. In sum, the received wisdom about folk psychology encapsulates two chief assumptions: (1) that making sense of actions requires interpreting them in terms of reasons composed of various propositional attitudes (at a bare minimum – beliefs and desires) and (2) that this activity is primarily concerned with providing predictions and explanations of actions. Folk psychology, construed in this way, is usually taken to be the central, core ability that underlies all interpersonal understanding and interaction, rather than just one amongst many ingredients of social ability. For instance, Currie and Sterelny (2000, p. 145) describe the orthodox view as being committed to the idea that "mind-reading [FP] and the capacity to negotiate the social world are not the same thing, but the former seems to be necessary for the latter [….] our basic grip on the social world depends on our being able to see our fellows as motivated by beliefs and desires we sometimes share and sometimes do not".

The central questions driving recent debates do not concern whether and where folk psychology is applied. Participants generally take the existence and ubiquity of folk psychology (construed as the attribution of internal propositional attitude states) for granted, concentrating instead on questions about which processes underlie it and how they arise during development. Great energy has been invested over the past two decades into determining the means by which folk psychology is conducted, the emphasis being on whether it involves the deployment of (1) a specialised theory, understood as a systematically organised body of knowledge detailing the links between typical perceptual inputs, intentional states and behaviours; (2) procedures of simulative imagining that directly manipulate the relevant intentional states themselves, without using any principles about such states (e.g. this might be achieved by using 'shared circuits' or by running practical reasoning and other sub-personal mechanisms off-line); or (3) some hybrid combination of these processes.

The name *theory* theory was first introduced by Morton (1980) in order to highlight the fact that the idea that folk psychology is a theory is itself a theory, and not obviously a true one.[2] Simulation theory, first advanced by Gordon in 1986 and subsequently developed by others, has been its most successful rival. A fairly typical simulationist claim is that, in order to predict and explain another person's psychological states and actions, one starts with an understanding of the target's current mental states and feeds pretend inputs into one's own mental state/behaviour generation mechanisms. Rather than using the output of this process to produce one's own actions, one assigns it to the other in the form of an action prediction. In other words, rather than applying a theory, one runs one's own psychological processes 'off-line', using oneself as a model of the other. However, there are many different versions of simulation theory and several points of disagreement between them.[3]

It was once common to think that there was a straight either/or choice between theory theory and simulation theory (cf. Stich and Nichols 1995). Currently, however, there is a growing trend towards the acceptance of some sort of hybrid account, where simulation routines and bodies of knowledge play complementary but distinct roles in interpersonal understanding. But what should be clear is that, for those who make this sort of debate the focus of their inquiries, the orthodox view of folk psychology is taken for granted. Theory, simulation and hybrid theories all presuppose a particular understanding of the nature, scope and function of our everyday modes of interpersonal understanding, which is that it consists of the attribution of propositional attitudes to others – encountered in the third person as a 'he', 'she' or 'it', rather than in the second-person as a 'you' – and perhaps also to oneself, in order to predict and explain action.

Such assumptions are equally left untouched in most debates about the biological basis, acquisition and development of folk psychological abilities. Many claim that folk psychological abilities have their source in an inherited device or 'module', an adaptation selected for the specific task of understanding others by facilitating simulation, deployment of a theory or both.[4] Others place more emphasis on the role of the child's developmental environment. For example, Garfield et al. (2001, p. 502) suggest that folk psychology is enabled by an "acquired module", which forms through the interaction of various in-built abilities with the social environment during development. Gopnik (1996) restricts the role of inherited components to a basic, non-metarepresentational starter theory of mind and rational theory-construction mechanisms, holding that folk psychology is a theoretical product fashioned in an evidence-sensitive way, one that is directly analogous to the way in which human adults forge mature scientific theories.

Another important topic that has been hotly debated is whether folk psychology is an exclusively human ability or one shared with other species. The most famous long-running and intensive debate about this has centred on the social intelligence of primates. Although it is not plausible that all sophisticated primate social intelligence requires a metarepresentational theory of mind, it was thought that some of their abilities might. For example, if primates were capable of genuine tactical deception, then they would need to be able to represent the beliefs, desires and intentions of others. In 1978, when the debate was just kicking off, Premack and Woodruff launched a small industry by asking in a paper of the same name, "Does the Chimpanzee have a theory of mind?"

Early assessments gleaned from anecdotes of the behaviours of individual animals seemed to show that a positive answer to this question might be warranted (Byrne and Whiten 1991). But more recent controlled experiments have decisively overturned that verdict. The dismal performance of chimpanzees on a non-verbal variant of the false belief task has galvanised widespread agreement that the social cognition of great apes does not depend on a sophisticated capacity for mindreading, certainly not one based on their having an understanding of belief or, indeed, of the inter-relations between that concept and the other central propositional attitudes. Comparing the test results for apes with those of human children has all but secured

the conviction that: "apes do not have a 'theory of mind' in the sense of understanding the false beliefs of others… [Thus such abilities] must have arisen in the human lineage only after human beings split from chimpanzees some 6–8 million years ago" (Call and Tomasello 1999, p. 393). Despite other deep and abiding disagreements, the two teams of researchers most closely involved in experimenting with chimpanzees agree on at least this much (Tomasello et al. 2003a, b; Povinelli and Vonk 2003; Povinelli and Vonk 2004).

The live issue today is not whether chimpanzees make use of a fully developed folk psychology but whether they have any degree of mindreading ability at all. Primatologists have, therefore, shifted the debate to a new level. The main question now occupying the field is: how do apes manage their sophisticated social interactions *without* the capacity for full-blown mentalistic understanding? What form does their social cognition actually take? What is important to note, yet again, is that, in making such cross-species comparisons, researchers have typically bought into the standard set of assumptions about the nature (propositional attitude attribution) and primary function (third-person prediction and explanation) of folk psychology.

These assumptions, although generally accepted, are debatable. Indeed, in recent years, there has been a small but growing murmur of dissent concerning the standard view of the nature and role of folk psychology, audible across the relevant disciplines. This dissent is now loud enough to suggest that the time has arrived for a reappraisal of received wisdom concerning the nature, role and scope of folk psychology. There is also the need to look at other aspects and perhaps other kinds of social understanding, which have been neglected as a consequence of the emphasis on belief-desire psychology and to consider how these might relate to belief-desire psychology.

The purpose of this volume is to bring together, for the first time, some of these dissenting voices in order to survey the various objections to the orthodox view and to explore proposals for re-orientating our understanding of folk psychology.

What could be wrong with the orthodox view? It is surely an open question as to whether our primary form of social cognition involves inferring the presence of hidden mental states. It could be that some aspects of mentality are readily perceivable in the expressive behaviour of others (Zahavi). Furthermore, perhaps the foundations of social understanding are better characterised in terms of *interactional* and embodied engagement with other people, rather than detached and *observational* abilities (Hobson). Closely associated is the possibility that folk psychology operates mostly in second-personal and not third-personal contexts. In other words, it is a matter of relatedness between an 'I' and a 'you', rather than of an 'I' observing a 'he', 'she' or 'it' (Stawarska). Such questions are important, given that legitimate worries can be raised about the utility of theory and simulation heuristics in contexts of second-person interaction. Indeed, there are powerful arguments for thinking that they are unnecessary for such social achievements (Gallagher).

If such concerns are justified, the idea that folk psychology is absolutely central to interpersonal social understanding stands in need of serious review. However, even if folk psychology is not the *basis* of most social understanding, perhaps it is

still the 'pinnacle' of human social achievements. In order to assess such a claim, it is important to consider what *kind* of understanding folk psychology engenders. Perhaps it is not, as is often claimed, theoretical but instead has an essentially narrative form. Furthermore, our facility with it might even stem from encounters with certain kinds of narratives (Hutto).

It is also debatable as to whether the primary function of folk psychology is to facilitate prediction and explanation. There are a number of other possibilities. For example, folk psychology might be essentially involved in the making of moral judgments (Knobe), enabling of social bonding (Andrews), and/or the moulding and regulating of behaviour (McGeer, Kusch). It might also be that folk psychology only plays a role in certain cultures, those with particular institutions and practices (Hutto, Kusch). More radically, could it even be that folk psychology is not a 'real' phenomenon? Perhaps the term groups together a wide range of disparate ways of effecting interpersonal understanding. If so, it may not reflect the way that people make sense of each other *at all*, being only a tidy philosophical schema for a much more complex set of phenomena, where propositional beliefs and desires might not feature in the way standardly supposed (Goldie, Morton, Ratcliffe).

Before turning to the task of outlining in more depth the structure of the volume and the nature of our contributors' various concerns about the orthodox view of folk psychology, it might be helpful to reflect on how that view came to predominate. This will make it easier to assess the strengths and weaknesses of the challenges and alternatives to it that are proposed in the following chapters.

1.2. THE IDEOLOGICAL BASIS OF THE RECEIVED VIEW

All the chapters in this volume are devoted to challenging the mainstream view of folk psychology in some way. For this reason, it is worth reminding the reader of certain salient, if familiar, facts about the ideological origins of received thinking on the topic. One, which we shall mention only en passant, is the long-standing tendency of analytic philosophers of mind to explicate intentional actions performed for reasons in terms of propositional attitudes, minimally beliefs and desires. This approach is deeply rooted and seldom challenged in philosophy of action. Anscombe's (1957) classic work, *Intention*, is often cited as the modern locus classicus, but the idea has a much more venerable history. It appears even in the account of purposeful acting provided by Aristotle, who tells us that "Intellect itself moves nothing….hence choice is either desiderative thought or intellectual desire" (*EN*, 1139a 35–36, 1139b 4–5). This explains the popularity of the tight definition of folk psychology that casts it as *necessarily* involving beliefs and desires, understood as propositional attitudes. If intentional action necessarily stems from the interaction between such states, it follows that to *understand* the basis of such actions, an understanding of what caused them (i.e. the reason for which they were done) is required.

It is a different matter to explicate why it is so widely assumed that the primary job of folk psychology is that of providing third-personal prediction and explanation. This can be achieved by looking back at some rather big movements in the recent

history of the philosophy of mind and cognitive science. What one finds is that the standard vision of folk psychology is structurally supported at several points. Essentially, a number of factors conspired to make theory theory the reigning view, leaving simulation to play the role of its natural, if relatively conservative, rival. And it was the rise of the idea that folk psychology is best understood as some kind of theory which fuelled the view that its core business was predictive-explanatory in character. But how did theory theory come to dominate?

Sellars is often regarded as being the first to have aired the idea that our understanding of mental states is, at root, theoretical. In his seminal 'Empiricism and the Philosophy of Mind'(1956/1997), he engaged in a bit of philosophical anthropological fiction. He famously mused about how our Rylean ancestors, who were as yet behaviourists, might have first fashioned an understanding of thoughts as inner episodes. He conjured up a mythical Jones who modelled inner thoughts on overt speech acts, imagining that the former could, just like the latter, be cited in the explanation of action. And if reasons are understood to be the inferred causes of action, it is nothing but a short step to thinking that *explaining* action using a schema that goes beyond what is merely given in another's outward responses is a kind of 'theoretical' activity. The theory theorist motto was and is: out of sight, into mind.[5]

An implication of supposing that genuinely intelligent engagements with the world are mediated by representations of it, of one sort or another, is that *representing* the representations of intelligent creatures requires taking a theoretical (or spectator's) stance towards them. In particular, it requires formulating hypotheses as to what exactly is 'going on in their minds' because this is not open to view. This was, of course, a decisive break from behaviourism, which sought to understand our everyday psychological concepts solely in terms of complex pairings of publicly observable stimuli and responses in a bid to ensure scientific credibility, as was wholly in line with the positivistic philosophy of science of the day.

Talk of 'unobservable' *episodes* later gave way to talk of 'abstract' or 'hypothetical' constructs, namely causally efficacious mental *states*. Mental states were thought to be entities that occupied causal roles, interacted with each other in complex ways and were identified in part by their typical causes and effects. From this thought about the nature of mental *states*, it is no great stretch to imagine that the meaning of mentalistic *concepts* might follow a similar pattern, being fixed by having appropriate links or relations to other concepts. That is, they might be defined by their place or role within a wider system of laws or, more softly, an inferential network. On this view, the very meaning of a particular mental concept is determined by the distinctive role it plays within a network of principles. In this respect, our familiar mentalistic vocabulary (i.e. our talk of thoughts, feelings and expectations) would be similar in important respects to other theoretically embedded vocabularies (i.e. talk of electrons, atoms and gravity) (Lewis 1970, 1978).[6]

That is a twice told tale. But recalling it helps to highlight the original sense in which mental terms were widely thought to be 'theoretical'. For, in philosophical circles, the popularity of theory theory was initially secured in large part by the

failure of empiricist theories of meaning, according to which mental terms were held to be grounded in ostensive definitions that referred either to introspected mental objects or to publicly observable behavioural dispositions. It is not an accident that Churchland relies so heavily on meaning holism to promote the idea that our commonsense understanding of minds, our propositional attitude psychology, is at root theoretical and open to the same kind of assessment as any other theory.

> If the meaning of our common observation terms is determined not by sensations, but by the network of common beliefs, assumptions, and principles in which they figure, then, barring some (surely insupportable) story about the incorrigibility of such beliefs, assumptions, and principles, our common observation terms are semantically embedded within a framework of sentences that has all the essential properties of a theory (Churchland 1979, p. 37).

Holism seems attractive because it makes it easier to see how concepts and categories need not be unalterably fixed. Conceptual schemes apparently develop and change over time. Our categories concerning 'what there is' are plastic, pliable and mutable. Indeed, it is precisely because our conceptions of the world shift and change that we can putatively make the sort of rare conceptual advances that constitute the progressive march of science and the growth of knowledge. Endorsing this kind of view, scientific theory theorists maintain that we start life with a basic theory of mind in place and that we actively develop it over the course of our childhood, modifying and forging concepts of mind in exactly the same way that scientists develop theories (Gopnik and Meltzoff 1997, p. 26). This, they hold, is the best way to account for the staged developments by which children gradually come by a mature theory of mind.

Nevertheless, one can be a theory theorist without buying into this view of the nature of mental concepts. Although theory theory is usually associated with the doctrine of meaning holism, many modern advocates of the idea that we operate with a 'theory of mind' are atomists. They hold that the core mentalistic concepts get their meaning denotationally by means of special mechanisms that 'lock on' to the relevant extensions. These concepts each play distinct roles in constituting the principles that make up the network of laws comprising one's theory of mind.

For example, Fodor maintains that the knowledge base of such domain-specific devices, including theory of mind mechanisms, have the "implicational structure of systems of semantically connected propositions" (1983, p. 7). To use his moniker, to accept this is to be committed to 'NeoCartesianism' about the content of such mechanisms. Indeed, the fact that they contain propositionally articulated principles, rules or schemas is precisely what fundamentally distinguishes them from psychological mechanisms of all other sorts. Not only do the individual principles have propositional contents; they must be bound together in coherent ways to form a theory. They have an internal *structure* appropriate to whatever is required to navigate the particular domain in question.

The unifying idea behind all of these proposals is that predicting how another creature might act (or, the flipside, to explain why it acted) requires representing its complex state of mind, in which certain propositional attitudes relate to one another in an appropriately structured way. In light of the developments discussed above,

despite minor disagreements about the meaning of mental terms and the dynamic character of such theories, it was pretty universally held that understanding others required having a *theory* of mind.

The corollary of this view is that the primary function of folk psychology is to provide predictions and explanations of actions, given what theories are used for elsewhere. It is because good theories are based on hard-won knowledge of a different sort than mere surface generalisations that they run deep. They provide powerful means of anticipating, explaining and controlling what happens, precisely because they tap into the world of the unseen and the abstract. A good theory will allow one to make solid bets that pay off quickly and selectively, based on minimal evidence. It will reliably guide expectations, even in novel circumstances.

But the question was bound to arise: Is folk psychology a *good* theory? The rest, as they say, is history. For, although propositional attitude psychology is generally thought to be a kind of theory of mind, no one supposed it to be the product of mature scientific theorising (even those who take it to be a product of the science-like theorising of children do not think it competes with the offerings of mature research on psychology). On the assumption that scientific psychology and folk psychology are both 'theories' in the same line of work, the latter looks like a poor contender when the two are compared. Without the benefit of a sustained and thriving research programme, it seems at best to be a *low-grade*, even stagnant, theory, which is in tension with the best theories of mind we can develop.[7]

Against this, some have held that folk psychology might be better understood as nothing more than a stance we adopt, or better a heuristic we employ, for the quick and dirty prediction of the behaviour of a wide variety of systems (Dennett 1985, 1987).[8] According to Dennett, we have no solid grounds for defending our folk psychological attributions above and beyond their success or otherwise in enabling us to make such predications. Closely associated with this is Dennett's view that folk psychological ascriptions are irredeemably indeterminate. In contrast, an ideal physics would allegedly provide perfect iron-clad predictions, in this way adjudicating decisively between rival interpretations. For this reason, the entities of such a science would be thought to describe fully objective, natural kinds. There would be, so to speak, objective 'matters of fact' when it comes to deciding between rival physical interpretations of 'what caused' the occurrence of certain other physical events. An ideal physics gets at these by trading in non-probabilistic laws that grant powers of perfect forward-facing prediction, and, conversely, perfect backward-looking explanations.[9] Putatively, the claim that 'physical kinds are the only real kinds' is not an arbitrary one. But if one accepts this kind of standard of reality, psychological phenomena turn out to be less than 'real'.[10]

Having mere instrumental value, the predicates of intentional psychology need not be in competition with the ontology of scientific naturalists. They can be treated as not more or less real than numbers, centres of gravity or other 'calculation-bound' entities.[11] Such crude predictive uses of folk psychology are, however, quite distinct from those in which we seemingly call on it to seriously *explain* actions. So this view comes at a heavy price. For it would seem, prima facie, that in citing

reasons we are attempting to get at the true causes of actions, and this does not sit easily with the idea that only science can tell us about the underlying mechanics of action.

Importantly, the verdict that folk psychology is 'inferior' to scientific psychology was seemingly secured not only by attending to the comparative accuracy of its predictions. For, it can be argued, the very fact that propositional attitudes are necessarily world-relating, in a way that takes us beyond the bounds of an organism's skin, reveals it to be wedded to an explanatory schema out of synch with the demands of a serious scientific psychology. The manner in which it assigns contents ultimately involves recourse to 'similarity conditions' and normative principles of charity that could not possibly feature in a mature science of the mind (Stich 1982; Davidson 1984, 1987). The ascription of propositional attitudes by means of radical interpretative methods, therefore, looks to be a kind of domestic anthropology, one bound up with a certain notion of 'rationality'. It is thus regarded as limited in scope and unfit to predict or explain the behaviour of exotic subjects, such as very young children, animals and confused or demented folk (Stich 1983).

Folk psychology was not only thought to be limited in this respect. It also broke faith with the alleged central tenet of any *bona fide* scientific psychology; that proper causal explanations should only cite proximate causes of action, those located physically, inside the skin of agents (Fodor 1981; McGinn 1989).[12] For some, the putative fact that beliefs and desires are individuated by factors that necessarily go beyond the skin meant that they could not be identified with inner states. This raised the concern that folk psychology could not provide causal explanations of action *at all*.[13]

All these considerations raised what was for some a threat and for others a hope: If folk psychology really was nothing more than a low-grade theory, perhaps the advances of cognitive science and neuroscience could offer us more genuine and fertile means of understanding ourselves. All sides acknowledged that the perceived stakes were high. But what is of great interest is that its friends and defenders alike hardly ever questioned folk psychology's 'theoretical' status, nor did they ever challenge the idea that its *primary* function was to provide third-personal predictions and explanations.[14] Such challenges would have clearly transformed the debate but never arose, perhaps because nearly everyone accepted that folk psychology's inferiority was to be excused for other reasons. For example, it would be *practically* impossible to come up with effective predictions and explanations in everyday life without making use of this schema, which was surely reliable enough and convenient. Its staunchest defenders, such as Fodor, held that for this reason folk psychology was a deep, powerful theory and ultimately the best way to understand others. Ironically, when push came to shove, its defenders emphasised its practical value rather than its purely theoretical virtues. But this line of argument presupposes that it is exceptionally good at yielding reliable third-personal predictions and explanations of action in everyday contexts.[15]

This swift review serves as a reminder not only of how the dual assumptions of the 'received view' took root; it also shows how deeply ingrained these are, at least for

those schooled in the analytic tradition. The trouble is that, despite being eminently challengeable, these twin assumptions have set agendas in a way that has hindered progress and constrained our imaginations. Of course, the arrival of simulation theory in 1986 gave us an alternative to the view that folk psychology depends upon a theory. However, simulation theories inherit much of the same historical philosophical baggage as theory theories. They take for granted the assumptions that (a) understanding people is most centrally a matter of attributing propositional attitudes in the third-person; and (b) its primary role is to predict and explain their behaviour. Thus, despite the many interesting discussions that have taken place concerning the relative merits of theory and simulation, the parameters of the theory-simulation debate are shaped by restrictive and questionable assumptions. Whilst these assumptions have held sway, important questions about folk psychology have simply not been raised and other ways of understanding it have not been explored. The papers collected here attempt to address this oversight.

1.2.1. Structure of the volume

The chapters in this volume appeal to work in a range of disciplines, including philosophy of mind, neuroscience and experimental psychology. A particularly conspicuous theme, running throughout the volume, is that phenomenology has a lot to offer when it comes to the study of interpersonal experience and understanding. Phenomenologists, including Husserl, Scheler and Merleau-Ponty, offer detailed and insightful descriptions of aspects of our interpersonal relations, which can be employed to challenge commonplace assumptions in the folk psychology literature. The relevance of phenomenology is made clear in Chapter 1 of Part I, 'Expression and Empathy', where Dan Zahavi takes issue with a presupposition common to all theory theories and some versions of simulation theory. According to these theories, detection of mental states involves an *inference to the best explanation* via which internal mental states are postulated on the basis of observed behaviour. Zahavi appeals to the writings of Scheler, Merleau-Ponty and others to argue that the connection between experience and expression is far closer than is usually maintained; one does not infer experience from expression but perceives it in expression. He challenges the assumption that mental states are wholly distinct from behaviour and argues that 'mere behaviour' is not something that we ordinarily experience. It is instead an abstraction from our experiences of others where perception of their expression *is* also perception of their experience. Zahavi further supports these phenomenological claims through an appeal to findings in developmental psychology. Theory and simulation, he concludes, are strategies that we might use in some circumstances but both are rare. Furthermore, they presuppose a more fundamental empathetic appreciation of others. 'Empathy', as conceived of by Zahavi, does not involve reaching out to a mind that is hidden behind another's behaviour. Instead, it is an immediate way in which we experience others; it is "to experience a behaviour as expressive of mind" (p. 37).

Zahavi's claims are supported and supplemented by Peter Hobson's account of the development of intersubjectivity in 'We Share, therefore We Think'. In this

chapter, Hobson suggests that both theory and simulation theories place too much emphasis on the interpretive abilities of socially isolated individuals. In contrast to such individualistic assumptions, he argues that social experience and understanding are founded upon patterns of affective relatedness between people. Like Zahavi, Hobson claims that we do not ordinarily experience and understand other people as 'minds plus bodies' but instead as whole persons, adding that our sense of *what people are* is constituted by reciprocal affective responsiveness. In support of this, he draws on several recent findings in experimental psychology, which suggest that social impairments in autism are largely due to diminished emotional relatedness to others. Hobson argues that, in the course of early typical development, affective interactions make an essential contribution to the development of social understanding. He proposes that most human beings, but perhaps not those with autism, have an innate propensity to 'identify with' other people. This identification does not usually take the form of wholly adopting another's perspective. Instead, we 'resonate' to the attitudes of others, whilst maintaining enough of our own perspective to yield the kinds of sharing that are a special feature of the social lives of human beings from infancy onwards. The meanings of expressions and gestures are, he suggests, perceived rather than inferred as internal causes of behaviour. Thus, Zahavi and Hobson agree that 'persons' are more basic than 'minds' and 'bodies', that we perceive the meaning of expressions and gestures, and that our sense of others is partly constituted by emotional responsiveness. Hobson adds to the picture a detailed developmental account that emphasises complex patterns of emotional interaction.

It might be argued that the emotional identification discussed by Zahavi and Hobson involves a kind of 'simulation' via which one employs one's own emotional abilities to model the emotional states of others. However, neither Zahavi nor Hobson thinks this is so and such simulationist claims are further disputed by Shaun Gallagher, in 'Logical and Phenomenological Arguments against Simulation Theory'. In this chapter, Gallagher distinguishes between theories of 'explicit' and 'implicit' simulation. According to the former, simulation is something that we *do* and of which we are aware. He challenges such theories on the grounds of phenomenological implausibility. For the most part, when we experience others as agents, we do so effortlessly and are not aware of modelling them or of undergoing an egocentric shift so as to adopt their perspectives. Gallagher acknowledges that this still leaves open the possibility of implicit simulation, according to which simulation is a sub-personal process that we are not ordinarily aware of employing. Several recent arguments for implicit simulation appeal to neurophysiological evidence concerning mirror neurons, which are cells that respond both when one acts in a certain way and when one perceives the same kind of action being performed by a conspecific. It has been suggested that mirror neurons amount to a mechanism via which the actions of others are modelled through the adoption of similar motor states. However, Gallagher argues that evidence from mirror neurons indicates that the actions, expressions and gestures of others are directly perceived rather than understood through a modelling process that follows perception of behaviour.

Mirror neurons are activated only 30–100 ms after visual stimulation. Their activity is directly elicited by perceptual stimuli, and there seems to be no place for a discrete modelling process. Hence, if we allow that perception is temporally extended and enactive (incorporating motor responsiveness), it seems that the mirror system is integral to perception. Gallagher, therefore, maintains that, for the most part, the meanings of gestures, expressions and actions are perceived in contexts of interaction between people. Theory and simulation are, at most, rare strategies that are deployed when behaviour is unusual and puzzling.

Zahavi, Hobson and Gallagher all distinguish second-person understandings of a 'you' or 'Thou' and third-person understandings of a 'she' or 'he'. Interestingly, no such distinction is made by participants in the theory-simulation debate, who assume that the structure of interpersonal understanding can be adequately conveyed in terms of the I–he/she/it relation. However, the significant differences between second- and third-person understanding are discussed by Stawarska in 'Persons, Pronouns and Perspectives'. She accepts, as her starting point, the kind of view proposed in the previous three chapters and goes on to address the question of how phenomenology might serve as a corrective for certain commonplace assumptions concerning folk psychology. Stawarska argues that both the folk psychology literature and the egological tradition in phenomenology suffer from a common failing, which is that they identify 'I' with 'the ego'. In doing so, they assume that the first-person pronoun behaves like a noun and inadvertently extricate it from its normal context of operation. In conjunction with this, they provide accounts of the ego that misleadingly isolate it from its relatedness to others. So in order to offer a phenomenological critique of folk psychology, one needs to look beyond the egological tradition. Stawarska then focuses on the distinction between the first, second and third person, arguing that the 'I-you' relation is importantly different from the relation between an 'I' and a 'he', 'she' or 'it'. The third-person pronoun can replace a noun phrase, such as 'the person next door'. However, 'I' and 'you' do not need to have pre-established referents; they are modes in which one encounters others, as addressor and addressee. Stawarska appeals to the work of Benveniste and others to argue that only 'I' and 'you' are properly 'personal', whereas referring to someone in the third person involves a withdrawal of a personal stance towards them. 'I' and 'you' arise together in dialogue and their reversibility is tied up with the reciprocity of communication. In contrast, to refer to someone as a 'he' or 'she' is to "ex-communicate" them, to remove them from the dialogical partnership. Stawarska also suggests that studies of natural language are telling with regard to our understanding of personal identity. Following Buber, she claims that 'I-Thou' comprises a unitary relation and that a grasp of this relation is closely tied to the development of an appreciation that others have perspectives or points of view. Indeed, to be able to adopt an 'I-Thou' stance just *is* to be open to the possibility of other people. This does not bode well for folk psychology, which draws exclusively on the I–he/she/it relation and, Stawarska thinks, unwittingly takes as its starting point a relation that amounts to withdrawal of the personal stance.

All four chapters in Part I indicate that there is a foundational appreciation of the personal that has been neglected by orthodox conceptions of folk psychology. This appreciation is closely tied to experiences that incorporate affective relatedness. Hence it is not something that can be adequately conveyed by an approach that emphasises a detached, observational standpoint as characteristic of interpersonal understanding. Now it is possible to agree with all this and yet still maintain that folk psychology has an important and distinctive role to play in social life. Experiencing others is not the same as offering explicit explanations of their actions and it is arguable that folk psychology, even if it does not comprise our most foundational sense of others as loci of experience and agency, does at least lie at the core of all explanations of action that appeal to reasons. This view is disputed in Chapter 1 of Part II, 'There are Reasons and Reasons', where Peter Goldie takes issue with the generally accepted assumption that beliefs and desires must play a central role in reason explanations. Goldie does not deny that all intentional actions *can* be explained in terms of beliefs and desires. However, he argues that reference to relevant beliefs and desires is seldom sufficient to provide a satisfactory explanation of action. Most everyday action explanations are, as he puts it, much "thicker" and also more varied than the kinds of belief-desire explanations that are offered as examples throughout the folk psychology literature. Goldie focuses on four additional factors that are often incorporated into action explanations:

1. Motives, such as revenge, which need not take the form of occurrent mental states
2. Character and personality traits, which dispose people to certain motives
3. Emotions, moods and undue influences, including anger, depression and drunkenness
4. Narrative-historical factors, which put behaviour in a broader context. For example, a narrative referring to someone's upbringing can do much to illuminate their actions.

Given these additional factors, he suggests that belief-desire explanations are often too skeletal to be adequately informative and that they are sometimes even redundant. He concludes that belief-desire psychology is only a part of everyday psychology, upon which philosophers and others have placed undue emphasis.

In 'Folk Psychology without Theory or Simulation', Daniel D. Hutto supports the more traditional idea that beliefs and desires, understood as interlaced propositional attitudes, are necessary for understanding intentional actions performed for reasons. When it comes to understanding a reason for action, this basic aspect of the folk psychological framework is always, at least, implied. Like other contributors, he denies that this sort of understanding is fundamental, even to basic human social coordination and understanding. Nevertheless, he argues that the kind of understanding that folk psychology affords is distinctive and vitally important to our lives and practices. But this is so precisely because its primary duty is not that of merely enabling us to make reliable third-personal predictions and explanations. Challenging the received wisdom on its views about the origins and applications of the folk psychological framework, Hutto promotes his 'Narrative Practice

Hypothesis' in place of theory theory and simulation accounts. He defends the view that it is through encounters with narratives about reasons for actions that humans come by an explicit understanding of and ability to use the folk psychological template as well as learning both how and when to apply it. By focusing on the features of such narratives, he argues that it is possible to account for our understanding of the structural features of folk psychology without understanding them or their source to be theoretical. Equally, he argues that his proposal accommodates the fact that the practical grasp of the various elements needed to gain a folk psychological understanding via narratives comes in stages. If so, this obviates the need to suppose that they are constructs of a developing theory. Against the claim that the capacity to understand the relevant kind of narrative must at least presuppose some kind of simulative ability, he concludes by admitting that, although some co-cognitive abilities are needed for acquiring and using the folk psychological framework, these do not amount to full-blown simulative abilities of the sort that presuppose a capacity to attribute reasons for action.

Victoria McGeer's contribution, 'The Regulative Dimension of Folk Psychology', recognises the ways in which we are deeply primed to develop a folk psychology, so strongly that we cannot help but apply it even to entities such as suitably animated geometric figures, which on reflection we would deny had any mentality at all. But what drives this strong tendency? Against the tradition, McGeer proposes that folk psychology does not involve detached use of a theoretical framework, employed primarily to effect predictions and explanations from an outsider's point of view. In place of this, she advocates a regulative conception of folk psychology, the core idea of which is that folk psychological competence is not exclusively or even primarily for the prediction and explanation of the actions of others; it is used to make ourselves intelligible and readable by others as much as for understanding them. In making folk psychological attributions, we are also laying down commitments as to how one ought to act and think, which must be abided by if the attributions are to hold good. In providing the basis for such normative constraints, McGeer argues that folk psychology acts as a kind of social glue, helping to mould and shape our actions, making possible a range of distinctive activities such as "cajoling, encouraging, reprimanding, promising…" (p. 149). By learning to partake in the kinds of shared norms that folk psychological understanding makes possible, we come to have a special sort of 'insider expertise', akin to the kind one has upon learning any other skill or competence. Thus, our ability to understand others in this way is the flipside of learning how to make ourselves understood by them. McGeer proposes, calling on an example from Ryle, that it is not unlike the competent chess player's ability to follow the games of others. She argues that early social interactions between children and caregivers, in which caregivers take the lead, are what foster this kind of mutual normatively governed readability. Shared understanding is thus rooted in a special kind of reciprocal attunement.

Also exploring the idea that 'folk psychology' may have functions other than the prediction and explanation of action, Joshua Knobe, in 'Folk Psychology: Science and Morals', concentrates on the relationship between belief-desire psychology and

moral judgments. He begins by noting that folk psychology clearly has a role to play in coming to moral judgments. For example, in blaming a person for something, we often presuppose that she did it intentionally. As attribution of intentions is part of folk psychology, it would seem that folk psychology generates the inputs for certain moral judgements. However, Knobe suggests that the relationship between folk psychology and moral considerations is not so tidy; moral considerations are not always *subsequent* to the attribution of folk psychological states. Indeed, he argues that they are *integral* to certain folk psychological concepts, such as intentions, reasons and values. Knobe provides new empirical evidence, indicating that the attribution of intentions, reasons and values is influenced by moral appraisals, and goes on to argue that the tendency to draw a clear distinction between folk psychology and ethics stems largely from the assumption that folk psychology is similar to a scientific theory. As moral considerations are not acceptable elements of scientific theories, they should also be kept separate from our folk psychological theory. However, he notes that not all theories are 'scientific'. If the term 'theory' is used permissively, we can also talk of religious theories, for instance, which certainly do incorporate ethical elements. Knobe adds that scientific theories have considerable predictive power but only at a price. They are 'special purpose' devices that apply to limited domains. One could postulate two separate special purpose theories for facilitating moral judgements and behaviour predictions. However, there would be significant overlap between their contents and the complexity of such an arrangement would make considerable, perhaps excessive, cognitive demands upon us. Knobe suggests instead that folk psychology is a more general tool, used both to predict behaviour and to make moral judgements. A single system that does both may not be as efficient as a dedicated system for one or the other but the pay-off in terms of cognitive economy is worth it.

The incorporation of ethical considerations into folk psychology is again suggestive of a regulative function, given that a *shared* ethical system can structure social interaction by specifying codes of conduct that all participants ought to abide by and are expected to abide by. We not only predict behaviour; we also criticise it, shape it and prescribe it. The regulative function of folk psychology is explored further by Martin Kusch, in *Folk Psychology and Freedom of the Will*. Kusch is critical of the orthodox conception of folk psychology for several reasons. First of all, it is overly restrictive, focusing primarily on beliefs, desires and actions, whilst ignoring other important folk concepts, including 'volition' and 'mood'. Such concepts are the topic of heated discussion elsewhere in philosophy but these discussions remain curiously cut off from the folk psychology debate. Of greater concern to Kusch is the individualistic emphasis in the folk psychology literature, which tends to focus on capacities possessed by individuals that allow them to predict and explain the actions of others. He suggests that more attention should be paid to the gregariousness and interdependence of people. We do not formulate our folk psychological intuitions in isolation from each other. Instead, intuitions are discussed, criticised and shared through communicative interaction. Kusch proposes that folk psychology is a shared framework for the interpretation of behaviour,

which regulates our conduct, rather than something to be found inside the heads of individuals. It is, he suggests, our "most fundamental social institution", a collective way that we take things to be, which is both descriptive and normative. Participation in this shared framework makes us predictable to each other and even shapes our experiences of our own mental states.

Kusch goes on to focus more specifically on the nature of 'free will', a central element of our everyday psychological discourse that is all but absent from recent discussions of folk psychology. He proposes a 'sociophilosophy' of free will, by which he means a philosophical view that acknowledges the extent of our sociability, and he does so through a critical discussion of Barnes' approach. Barnes makes a distinction between pragmatic and strict criteria for an action's being the result of free choice. According to the strict, metaphysical criterion, it must have originated from an 'uncaused causer'. However, everyday discussion in legal and other contexts accepts that voluntary actions are causally influenced by all manner of factors and yet still draws a distinction between those actions we perform freely and those we do not. Central to everyday distinctions is the difference between decisions that can be influenced by the verbal intervention of another person and those that cannot. But what of the strict criterion? When we influence people's decisions, we often do so by appealing to their rational nature, to their freedom and responsibility. So appeals to free choice, in the strict sense, are employed as part of the social practice of regulating the conduct of others. On the basis of this, Barnes offers a compatibilist account of free will. Kusch, however, raises the concern that this account slides first into the view that free will is a necessary illusion and then on into eliminativism. He revises it to offer a constructivist account, according to which free will, like belief and other folk psychological categories, is something that people have in virtue of their participation in a contingent social institution. The phenomenology of free will is itself shaped by that institution. This, Kusch suggests, does allow for a form of compatibilism.

All four chapters in Part II question, in different ways, the nature, breadth and function of folk psychology, suggesting that orthodox conceptions are too restrictive in their emphasis on beliefs, desires and actions. Many other factors play a role in our predictions and explanations of others. Furthermore, prediction and explanation are not the sole functions of folk psychology. It also supplies shared norms of conduct that contribute to interpersonal understanding, interaction and coordination. These norms need not be understood as contents of individual brains, to be interpreted by other individual brains. Instead, they might well take the form of a social institution, through which we interpret and experience ourselves and others.

The chapters in Part III emphasise the heterogeneity of folk psychology and the final two even question whether there is such a thing as 'folk psychology'. This part of the book begins with a discussion of 'Critter psychology', by Kristin Andrews. Andrews assumes that humans do have a folk psychology and focuses on the question of whether certain other animal species also have their own folk psychology. In order to address this question, she first asks what is meant by 'folk psychology' and notes that orthodox conceptions are rather narrow in

scope, emphasising the attribution of propositional attitudes to explain and predict behaviour as the ability that is central to human social life. Andrews challenges this narrow conception, noting that we don't just predict and explain people's behaviour; we coordinate, interact and bond with them and we also sometimes try to deceive them. The abilities underlying these achievements are, she suggests, quite heterogeneous in nature and thus 'folk psychology', construed narrowly as 'belief-desire' psychology, actually has a very limited role. Andrews adopts a more liberal conception, which accommodates the full range of our diverse social achievements and, given this conception, asks whether certain other species also have a folk psychology. The answer, she says, is yes, as many other animals display at least some social abilities. Andrews then addresses the more specific question of whether certain other species possess an appreciation of 'mindedness' and 'intention', which, she argues, need not presuppose an ability to assign beliefs. She focuses on studies of chimpanzee social behaviour and argues that they are sufficient to demonstrate that chimpanzees can conceptualise 'seeing', a mentalistic concept. However, chimpanzee 'seeing' differs from the human concept of 'seeing'; the latter is closely associated with 'believing', whilst the former may not be. So it seems that, if one places less emphasis on an understanding of beliefs and desires, the question of whether non-human animals have a folk psychology cannot be answered with a simple yes or no. It is instead a question of which folk psychological abilities are possessed by which species.

The theme of folk psychology's heterogeneity is pushed further by Adam Morton, in 'Folk Psychology does not Exist'. In this chapter, Morton considers the possibility that the label 'folk psychology' refers to a diverse bundle of abilities and that a unitary account of our capacity for folk psychology will, therefore, be unavailable. He notes that we find ourselves in various different kinds of social situation, where we need quite different kinds of information. Many of the diverse skills that we apply in order to interpret and interact with others will not even be specific to interpersonal understanding. For example, self-preservation skills are brought to bear on people, animals and natural disasters, all of which can be dangerous. Social coordination also draws on an explicit understanding of strategic choice, which plays an important role in one's own decision-making and is therefore not specific to interpreting other people. Morton also considers the possibility that there will be cultural differences in how a range of capacities are applied to understand other people and, agreeing with McGeer and Kusch, remarks that a culture's folk conception of interpersonal understanding plays a role in organising the way people think about themselves and others, thus making them predictable to each other. He goes on to suggest that the categories 'belief' and 'desire' are ambiguous and that the kinds of states we attribute when we use such terms are made clearer by contexts of utterance and other factors. Morton concludes with the suggestion that the label 'folk psychology' groups together a range of overlapping capacities that could just as well be grouped in quite different ways. As with the classification of constellations, like the Great Bear, folk psychology does not pick out a unitary ability but imposes order upon a disparate bundle of abilities. That folk psychology

is 'as real' as a constellation suggests that it has no *psychological* reality and exists only as a contingent way of ordering and systematising social life, in a manner similar to that suggested by Kusch.

In 'From Folk Psychology to Commonsense', Ratcliffe also suggests that there is no such thing as folk psychology. However, he takes a different approach to Morton and considers what is actually meant by the terms 'folk' and 'commonsense'. On the basis of a study involving questionnaires, Ratcliffe argues that belief-desire psychology is not 'commonsensical', meaning that it is not something that people articulate when asked what interpersonal understanding consists of. Neither is a 'commonsense belief-desire psychology' something discovered via empirical science, given that empirical studies tend to presuppose it. Ratcliffe concludes that the orthodox conception of folk psychology is in fact a philosophical position, whose nature as such is obscured by the labels 'folk' and 'commonsense'. Articulating our most fundamental conceptual commitments is a difficult philosophical task and Ratcliffe argues that proponents of a belief-desire psychology have not done the required philosophical work. He also argues that the *scope* claimed for folk psychology is quite unclear. Ratcliffe then addresses the *nature* of folk psychology, focusing on the concepts of 'belief' and 'desire'. Like Morton, he argues that these concepts are ambiguous to say the least. In fact, they are abstract placeholders for a wide range of states that we distinguish with ease in everyday life. Ratcliffe suggests that belief-desire psychology is not, strictly speaking, false. Instead, it is a vague abstraction from social life that has no psychological reality and is not something that the so-called folk actually employ. His position complements the views of Morton and Goldie in certain respects. However, unlike Goldie, he argues that 'belief' and 'desire' do not have even a minimal role to play in action explanation. This is because they are, as Needham (1972) put it, "peg words" for a range of different states that we appeal to in our explanations of action.

1.2.2. The future of folk psychological research

The contributors to this volume are all united in the view that questions concerning the existence, nature, scope and purpose of 'folk psychology' have been insufficiently explored and that presuppositions which have shaped the theory-simulation debate need to be made explicit and questioned. Although the positive accounts offered here differ in various ways, together they not only comprise a persuasive critique of the orthodox conception of folk psychology but also point to new ways of understanding it. A view consistently promoted in the following chapters is that, in most of our engagements with others, we do not ordinarily infer the presence of hidden mental states but perceive and emotionally respond to embodied expressions of mentality. A great deal of social understanding occurs in contexts in which we interact with others, rather than being a matter of detached observation. And it is through such perceptually and emotionally guided engagements that our most basic sense of what is unique about other persons is first grasped.

If this is indeed the case, belief-desire psychology is neither ubiquitous nor fundamental. Indeed, a common theme throughout this volume is that it is frequently

over-emphasised and must be supplemented by various other abilities and forms of social understanding. This is accompanied by the concern that the nature of its role in our lives and practices needs to be reconsidered.

In summary, all the contributors to this volume acknowledge that 'folk psychology', construed as belief-desire psychology, is limited in scope and needs to be supplemented and/or differently characterised in some way. However, it should not be assumed that they all agree on precisely how it should be revised. Many chapters do complement each other but there are several points of disagreement concerning the nature, role and scope of folk psychology. For example, not all agree that there is even such practice as 'belief-desire psychology' or that the term 'folk psychology' has any useful role to play. These and the many other issues raised here suggest that recent debates concerning folk psychology, theory theory and simulation theory have been too restricted in scope and that a more encompassing approach to the topic of interpersonal understanding, interaction and coordination is required.

ACKNOWLEDGEMENTS

We are grateful to all the contributors to this volume for reading an earlier version of this introduction and providing us with helpful feedback.

NOTES

[1] Some authors, such as Churchland (1981), argue that there will be no place for propositional attitudes in a mature cognitive science. Others, including Fodor (1987) vehemently disagree, holding that they will feature in our best science of the mind. Others hold that, although reasons do in fact cause actions, there is no prospect of folk psychology ever becoming a law-like special science (cf. Davidson 1980).

[2] Morton (1980, p. 170) was perhaps the first to promote the idea that we understand folk psychology not as a 'theory' *per se* but as being, at least in part, a skill-based practice.

[3] See, for example, the contrast between Gordon (1996) and Heal (1996).

[4] For examples of such accounts, see Fodor (1995), Leslie and German (1995), Carruthers (1996), Segal (1996) and Baron-Cohen and Swettenham (1996), all of whom favour versions of the theory theory or a theory/simulation mix. Gordon (1995) suggests that there might be a module for the 'practical simulation' of others.

[5] Strictly speaking, having the property of being unobservable is not a requirement for something to be theoretical. It is possible to make visible what was once only theoretical, as has been achieved with DNA, electrons and the like. What is important is that such definitions of theoretical terms are 'not simply restatements of the data' (Gopnik and Meltzoff 1997, p. 35, see also Wellman 1991, p. 20).

[6] In fact, endorsing this view of the meaning of mental predicates is not sufficient to underwrite the idea that such terms are theoretical. To accept that the content of our mental concepts is constituted holistically at best establishes that they are similar to theoretical terms (to the extent that both share this characteristic). Theoretical concepts could be a sub-set of the kind of concepts that gain their meaning holistically. We cannot infer that mentalistic concepts are theoretical even if we assume that they only have their meaning by being part of an interrelated network of concepts. They may be holistic without being theoretical per se and even without being theory-like in any other respect.

[7] Even Davidson subscribed to this view, despite maintaining that folk psychological explanations are irreducible to physical ones because of "the normative character of mental concepts" (Davidson 1987, p. 46). The idea is roughly this: We cannot assign length without a physical framework. Similarly,

we cannot ascribe propositional attitudes without a normative, interpretative framework. Each of these domains has its own constitutive features. As long as we conceive of people 'as rational', we cannot have a closed system for propositional ascriptions, such that it could be suitably reduced to a system of mere physical descriptions. Making such ascriptions is, for this reason, irredeemably disanalogous to the way in which we understand the behaviour of 'mindless' entities (Davidson 1991, pp. 162–163, see also Davidson 1996). In this light, it is both surprising and somewhat misleading for Davidson to say that folk psychology is a "familiar mode of explanation of human behaviour and must be considered an organised department of common sense which may as well be called a theory" (Davidson 1984, p. 158). In any case, like straightforward theory theorists such as Fodor, Davidson maintains that propositional attitude psychology has special constitutive features that make it unique amongst 'theories'.

[8] On this view "the pattern of belief is discernible in an agent's (observable) behaviour when we subject it to 'radical interpretation'" (Dennett 1991a, p. 30). Of interest, however, is the fact that Dennett held that "a good reason for not calling [intentional psychology] a theory, [is that] it does not consist of any explicit theorems or laws" (Dennett 1991b, p. 134).

[9] This kind of thinking was widespread and had deep hooks. For example, although Davidson differed from Dennett in thinking that folk psychology could provide genuine explanations of actions (at least of the singular causal variety), he accepted that these were nevertheless inferior when compared to the predictions and explanations that – in principle – would ensue from an ideal physics. Although he insisted that ideal physical explanations could never replace those of folk psychology in the way they could replace those of folk physics, he was inclined to make remarks suggesting, nonetheless, that the absence of folk psychological laws revealed that "reason-explanations, by virtue of the features I have been depicting, are in some sense low-grade; they *explain less* than the best explanations in the hard sciences because of their heavy dependence on causal propensities" (Davidson 1987, p. 42, emphasis added, cf. also Davidson 1990, p. 7). There is however reason to think that Davidson's line on this is unstable (see Hutto 1998, 1999).

[10] We have good independent reasons to question this standard of what is to count as a natural kind. If we did believe that only ideal physical kinds were real, (a) we would have to treat the categories of nearly every other discourse – including the non-reducibly special sciences – less than seriously; and (b) we may doubt that there are any real kinds at all because we might doubt that an ideal physics is a genuine possibility (as is the view of some exclusive scientific realists, see Churchland 1989, pp. 294–295).

[11] Officially, Dennett is not an eliminativist about beliefs and desires – and although his position is often regarded as a form of instrumentalism, he prefers to be thought of as a *mild realist* (cf. Dennett, 1991a). On his account, we are invited to think of the 'posits' of intentional psychology as species of *abstracta* and not as serious theoretical posits (i.e. *illata*). An example of *abstracta* would be Dennett's lost sock centre, defined as "the center of the smallest sphere that can be inscribed around all the socks I have ever lost in my life" (Dennett 1991a, p. 28). In this way the 'posits' of folk psychology are to be thought of as akin to numbers and other abstract objects.

[12] Fodor, like Stich, accepts that any scientifically respectable explanation must be of a causal character. Essentially, his methodological solipsism plus his commitment to physicalism entails acceptance of a version of Stich's principle of autonomy. The principle of autonomy states "psychological states and relations, or at least those that should be of concern in a serious scientific psychology, must supervene on the current, internal, physical states of the organism(s) in question" (Stich 1982, p. 197). Supervenience (in some varieties) is thought to provide a means of describing the nature of a non-reductive relationship of dependency between two types of properties, aspects or predicates. Fodor denies that Stich's principle of autonomy is identical to his formality condition (cf. Fodor 1987, p. 42); but, ultimately, the problem for the supervenience of contentful properties on internal states emerges on either formulation.

[13] Whether it could or not rested on its prospects for a clean reduction to a more basic science – or at least the prospects for being regarded no worse in this respect than any other 'special' science. Arguing for the latter was Fodor's main stratagem to secure folk psychology's credentials (see Fodor 1987).

[14] In her book *Saving Belief*, Lynne Rudder Baker recognises that the elimination of commonsense psychology is akin to a kind of cognitive suicide because of its role in underwriting a wide variety of key practices. She offers the following list: (i) social practices that depend upon ordinary explanation

and prediction of behaviour would become unintelligible; (ii) moral and legal practices would become senseless; (iii) linguistic practices would become mysterious; and (iv) psychological theorising would become problematic (Baker 1988, pp. 130–133). While this catalogue points to folk psychology's indispensability, it also highlights that folk psychology may have more vital roles in our lives than the mere prediction and explanation of behaviour.

[15] While this is frequently said, on close inspection, there are reasons to think that it is not so (Hutto 2004).

REFERENCES

Anscombe GEM (1957) Intention. Blackwell, Oxford.

Aristotle (1984) The complete works of Aristotle: the revised Oxford translation. Princeton University Press, Princeton, NJ.

Baker LR (1988) Saving belief: a critique of physicalism. Princeton University Press, New Jersey.

Baron-Cohen S, Swettenham J (1996) The relationship between SAM and ToMM: two hypotheses. In Carruthers P, Smith PK (eds): pp 158–168.

Byrne RW, Whiten A (1991) Computation and mindreading in primate tactical deception. In: Whiten A (ed) Natural theories of mind. Blackwell, Oxford. pp 127–142.

Call J, Tomasello M (1999) A nonverbal false belief task: the performance of children and great apes. Child Dev 70: 381–395.

Carruthers P (1996) Simulation and self-knowledge: A defence of the theory-theory. In: Carruthers P, Smith PK (eds), pp 22–38.

Churchland PM (1979) Scientific realism and the plasticity of mind. Cambridge University Press, Cambridge.

Churchland PM (1981) Eliminative materialism and the propositional attitudes. J Philos 78:67–90.

Churchland PM (1989) A neurocomputational perspective: the nature of mind and the structure of science. MIT Press, Cambridge, MA.

Currie G, Sterelny K (2000) How to think about the modularity of mind-reading. Philos Q 50:145–160.

Davidson D (1980) Essays on actions and events. Clarendon Press, Oxford.

Davidson D (1984) Inquiries into truth and interpretation. Clarendon Press, Oxford.

Davidson D (1987) Problems in the explanation of action. In: Pettit P, Sylvan R, Norman J (eds) Metaphysics and morality. Blackwell, Oxford pp 35–49.

Davidson D (1990) The structure and content of truth. J Philos LXXXVIII.

Davidson D (1991) Three Varieties of Knowledge. In: Phillips Giffiths A (ed) A.J. Ayer Memorial essays. Cambridge University Press, Cambridge, pp 153–166.

Davidson D (1996) Subjective, intersubjective, objective. In: Coates P, Hutto DD (eds) Current issues in idealism. Thoemmes Press, Bristol.

Davies M, Stone T (eds) (1995a) Mental simulation: evaluations and applications. Blackwell, Oxford.

Davies M, Stone T (eds) (1995b) Folk psychology: the theory of mind debate. Blackwell, Oxford.

Dennett DC (1985) Brainstorms. MIT Press, Cambridge, MA.

Dennett DC (1987) The intentional stance. MIT Press, Cambridge, MA.

Dennett DC (1991a) Real patterns. J Philos: 27–51.

Dennett DC (1991b) Granny's campaign for safe science. In: Loewer Barry (ed) Meaning in mind: Fodor and his critics. Blackwell, Cambridge pp 87–94.

Fodor JA (1981) Representations: philosophical essays on the foundations of cognitive science. MIT Press, Cambridge, MA.

Fodor JA (1983) The modularity of mind. MIT Press, Cambridge, MA.

Fodor JA (1987) Psychosemantics. MIT Press, Cambridge, MA.

Fodor JA (1995a) A theory of the child's theory of mind. In: Davies M, Stone T (eds), pp 109–122.

Garfield J, Peterson C, Perry T (2001) Social cognition, language acquisition and the development of the theory of mind. Mind Lang 16: 494–541.

Gopnik A (1996) Theories and modules: creation myths, developmental realities, and Neurath's boat. In: Carruthers P, Smith PK (eds), pp 169–183.
Gopnik A, Meltzoff AN (1997) Words, thoughts, and theories. MIT Press, Cambridge, MA.
Gordon R (1995) Folk psychology as simulation. In: Davies M, Stone T (eds) (a), pp 60–73.
Gordon R (1996) 'Radical' simulation. In: Carruthers P, Smith PK (eds), pp 11–21.
Heal J (1996) Simulation, theory and content. In: Carruthers P, Smith PK (eds), pp 75–89.
Hornsby J (1987) Simple mindedness: in defense of naive naturalism in the Philosophy of mind. Harvard, University Press, Cambridge, MA.
Hutto DD (1998) Davidson's identity crisis. Dialectica 52(1): 45–61.
Hutto DD (1999) A cause for concern: reasons, causes and explanations. Philos Phenom Res 59(2): 381–401.
Hutto DD (2004) The limits of spectatorial folk psychology. Mind Lang 19:548–573.
Leslie AM, German TP (1995a) Knowledge and ability in 'Theory of Mind': a one-eyed overview of the debate. In: Davies M, Stone T (eds), pp 123–150.
Lewis D (1970) How to define theoretical terms. J Philos 67:427–446.
Lewis D (1978) Psychophysical and theoretical identifications. Australas J Philos 50:249–558.
McGinn C (1989) Mental content. Basil Blackwell, Oxford.
Morton A (1980) Frames of mind: constraints on the common sense conception of the mental. Oxford University Press, Oxford.
Needham R (1972) Belief, language, and experience. Blackwell, Oxford.
Povinelli DJ, Vonk J (2003) Chimpanzee minds: suspiciously human? Trends Cogn Sci 7:157–160.
Povinelli DJ, Vonk J (2004) We don't need a microscope to explore the chimpanzee's mind. Mind Lang 19:1–28.
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci:515–526.
Ratcliffe M (2005) Folk psychology and the biological basis of intersubjectivity. R Inst Philos Suppl 56: Philos Biol Life:211–233.
Segal G (1996) The modularity of theory of mind. In: Carruthers P, Smith PK (eds), pp 141–157.
Sellars W (1956/1997) Empiricism and the philosophy of mind. Harvard University Press, Cambridge, MA.
Stich S (1982) On the ascription of content. In: Woodfield A (ed) Thought and Object. Oxford University Press, Oxford pp 153–206.
Stich S (1983) From folk psychology to cognitive science. MIT Press, Cambridge, MA.
Stich S, Nichols S (1995) Folk psychology: simulation or tacit theory? In: Davies M, Stone T (eds) (a), pp 123–158.
Tomasello M, Call J, Hare B (2003a) Chimpanzees understand psychological states – the question is which ones and to what extent. Trends Cogn Sci 7:153–156.
Tomasello M, Call J, Hare B (2003b) Chimpanzees versus humans: it's not that simple. Trends Cogn Sci 7:239–240.
Wellman HM (1991) From desires to belief: acquisition of a theory of mind. In: Whiten A (ed) Natural theories of mind. Blackwell, Oxford pp 19–38.

PART I

EMOTION, PERCEPTION, AND INTERACTION

2. EXPRESSION AND EMPATHY

The ongoing debate about the nature of social cognition has been dominated by two competing positions, the theory theory of mind and the simulation theory of mind. Although these positions are regularly depicted as being quite divergent, I will in the following discuss what I take to be a shared assumption, namely a certain conception of the mind-body relation. I will criticize it and, drawing on thinkers like Scheler, Merleau-Ponty and Wittgenstein, I will argue that our understanding of others is crucially dependent on our understanding of their expressive behaviour.

2.1. THE THEORY OF MIND DEBATE

In recent years, much of the discussion of the nature of social cognition has taken place within the framework of the so-called theory of mind debate. The expression "theory of mind" is generally used as shorthand for our ability to attribute mental states to self and others and to interpret, predict, and explain behaviour in terms of mental states such as intentions, beliefs and desires (cf. Premack and Woodruff 1978, p. 515). Although it was originally assumed that it was the possession and use of a *theory* that provided the individual with the capacity to attribute mental states, the contemporary debate is split on the issue and is generally considered to be a dispute between two views. On one side, we find the *theory theory of mind* and on the other the *simulation theory of mind*.

This neat division is an oversimplification, however. Not only because of the existence of several hybrid theories, but also because neither of the main positions are theoretical monoliths. Theory theorists are basically split on the issue of whether the theory in question is innate and modularized (Carruthers, Baron-Cohen) or whether it is acquired in the same manner as ordinary scientific theories (Gopnik, Wellman). As for the simulationists, some claim that the simulation in question involves the exercise of conscious imagination and deliberative inference (Goldman), some insist that the simulation although explicit is non-inferential in nature (Gordon, 1996), and finally there are those who argue that the simulation rather than being explicit and conscious is implicit and subpersonal (Gallese) (cf. Gallagher 2007).

Generally speaking, however, the theory theory holds that the understanding of minded beings (be it oneself or others) is theoretical, inferential and quasi-scientific in nature. It views the attribution of mental states as a question of an inference to the best explanation and prediction of behavioural data and argues that mental states are unobservable and theoretically postulated entities. It consequently denies that we have any direct experience of such states. Many philosophers (phenomenologists

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 25–40.

included) would claim that we need concepts in order to extract and comprehend the informational richness of what is already given, already present to us (like the connoisseur who is able to discern and differentiate aromas and flavours in the wine to which others are not sensitive). Many would also endorse the idea that our observations are influenced and enriched by former experiences. When the theory theory claims that the attribution of mental states is theoretically mediated, it has something more radical in mind. The idea is basically that the employment of theory allows us to transcend what is given in experience:

> One of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others (and indeed of ourselves) are completely hidden from the senses, they can only ever be inferred (Leslie 1987, p. 139).
>
> Normal humans everywhere not only "paint" their world with color, they also "paint" beliefs, intentions, feelings, hopes, desires, and pretenses onto agents in their social world. They do this despite the fact that no human has ever seen a thought, a belief, or an intention (Tooby and Cosmides in Baron-Cohen 1995, p. xvii).

It should be noticed that the theory theory of mind defends a double thesis. It does not only claim that our understanding of others is inferential in nature, it also argues that our own self-experience is theoretically mediated. After all, the basic idea is that *any* reference to mental states involves a theoretical stance, involves the application of a theory of mind.

Whereas the theory theory argues that our understanding of others chiefly engages detached intellectual processes, moving by inference from one belief to the other, the simulation theory of mind argues that our understanding of others exploits our own motivational and emotional resources. Thus, in contrast to the theory theorists, the simulationists would deny that what lies at the root of our mentalizing abilities is a sort of theory. In their view, we possess no such theory, or at least none complete enough to underpin all our competence with psychological notions. This far, the different versions of the simulation theory agree. However, when it comes to a more positive account of what the simulationist alternative amounts to, the opinions differ. Given restrictions of space, I will in the following focus on Goldman, and every time I refer to the simulation theory, I will have his theory in mind. My reason for this focus is quite simple. I think Goldman's version of the simulation theory is the one that most unequivocally relies on and refers to a routine that merits the name "simulation".[1]

According to Goldman, we don't need a theory in order to understand others. Rather, we can simply use our own minds as a model. Our understanding of the minds of others would be grounded in our introspective access to our own mind; our capacity for self-ascription precedes the capacity for other-ascription. More specifically, Goldman argues that my understanding of others is rooted in my ability to project myself imaginatively into their situation. I literally use my imagination to put myself in the target's "mental shoes". If I for instance witness an immigrant being harassed by a desk clerk, I would be able to grasp the immigrant's mental state and predict his subsequent behaviour by means of the following procedure. By means of an explicit simulation, I would imaginatively put myself in his situation,

I would imagine how I would feel and react under similar circumstances and on the basis of analogy I would then attribute or project similar states to the person I am simulating (cf. Goldman 2000).

In my view, both sides in the theory of mind debate are faced with difficulties. When it comes to the simulation theory of mind, one might initially question whether there is any experiential evidence in support of the claim that our understanding of others relies on conscious simulation routines. As Wittgenstein once remarked, "Do you look into *yourself* in order to recognize the fury in *his* face?" (Wittgenstein 1981, Sect. 220). Furthermore, one might ask whether it is really legitimate to cast our experience of others in terms of an imaginative exercise. When we project ourselves imaginatively into the perspective of the other, when we put ourselves in his or her shoes, will we then really attain an understanding of the other or will we merely be reiterating ourselves? To put it differently, will a process of simulation ever allow for a true understanding of the *other* or will it merely let me attain an understanding of myself in a different situation? As for the theory theory of mind, one could for instance question some of its empirical claims and implications. If a theory of mind is required for the experience of minded beings, then any creature that lacks such a theory will also lack both self-experience and experience of others. According to the standard view, however, children only gain possession of a theory of mind when they are around four years old. Thus, a direct implication of the theory theory of mind seems to be that young children will lack any understanding of self and other during the first three to four years of life. But is that really true? When pressed on the issue, some theory theorists refer to various mechanisms that might be regarded as precursors to a theory of mind (cf. Baron-Cohen 1995) and will in fact concede that children do understand (experience) psychological states such as emotions, perceptions and desires in both self and other prior to the possession of a proper theory of mind. They then argue that what these children lack is an understanding of *representational* mental states (cf. Wellman et al. 2001, pp. 656, 677). However, since the term "representational mental state" is quite ambiguous, this admission doesn't do much to clarify the situation. At times, the term is used inclusively to cover all intentional states, including perceptions; at other times, it is used much more restrictively to cover only proper beliefs (thoughts). This vacillation makes the theory theory into something of a moving target. It also threatens to leave it with the uncomfortable choice between only two options. It can either defend a very strong, some would say extreme, claim, according to which the child has no first-person access to any of its own mental episodes and no experience of other minded creatures prior to the acquisition of a theory of mind. It can, alternately, defend a much weaker, some would say trivial, claim by defining representational mental states in such narrow terms that it is no wonder that it takes a relatively high level of cognitive sophistication to be able to understand and attribute them to self and other. To rephrase the criticism in slightly different terms: One can define a mental state as something purely interior and private, as something that is not visible in meaningful actions and expressive behaviour. Given such a concept of a mental state, there are good reasons to believe that children will only be able

to master the concept and ascribe it to others and self at a relatively late stage. However, the obvious and crucial question is why one would want to operate with such a narrow mentalistic understanding of the mind in the first place.[2]

My aim in the following will not to be to pursue and develop these lines of criticism further. Rather I wish to focus on a slightly different issue. The theory theory of mind and the simulation theory of mind are frequently depicted as quite opposed accounts of the basic nature of social cognition. However, both accounts share certain presuppositions that underlie and shape the very theory of mind debate. In particular, they both share certain assumptions about the mind-body relation. I find these assumptions questionable, and what I intend to do in the following is to suggest that an alternative and more satisfactory account can be found in phenomenology.

2.2. THE ARGUMENT FROM ANALOGY

Let me take my point of departure in the classical attempt to come to grips with the problem of other minds known as the *argument from analogy*. The argument runs as follows: The only mind I have direct access to is my own. My access to the mind of another is always mediated by his bodily behaviour. But how can the perception of another person's body provide me with information about his mind? In my own case, I can observe that I have experiences when my body is causally influenced, and that these experiences frequently bring about certain actions. I observe that other bodies are influenced and act in similar manners, and I therefore infer per analogy that the behaviour of foreign bodies is associated with experiences similar to those I have myself. In my own case, being scolded by hot water is associated with the feeling of intense pain, this experience then gives rise to a quite distinct behaviour: screaming. When I observe other bodies being scolded by hot water and screaming, I infer that it is likely that they are also feeling pain. Thus, the argument from analogy can be interpreted as an inference to the best explanation. An inference bringing us from observed public behaviour to a hidden mental cause. Although this inference does not provide me with indubitable knowledge about others and although it does not allow me to actually experience other minds, at least it gives me more reason to believe in their existence, than in denying it.

What is the relationship between the classical argument from analogy and the contemporary theory of mind debate? More specifically, how does the argument from analogy relate to the simulation theory and to the theory theory respectively? I think the argument straddles the difference between the two alternatives. It has affinities with (one version of) the simulation theory in so far as it argues that we have an immediate and direct access to the content of our own minds, and in so far as it holds that this self-acquaintance serves as our point of departure when it comes to an understanding of others, that is, in so far as it insists that we come to know others in analogy with ourselves. On the other hand, by arguing that our understanding of others is an inference to the best explanation, an inference bringing

us from observed public behaviour to a hidden mental cause, the argument from analogy also shares fundamental claims with the theory theory of mind.

Needless to say, since the theory of mind debate is a rather recent phenomenon, none of the classical phenomenologists have criticized it *expressis verbis*. But the phenomenologists have indeed questioned the validity of the argument from analogy. In his book *Wesen und Formen der Sympathie*, for instance, Scheler subjects the argument to a scathing criticism. As he points out, the argument presupposes that which it is meant to explain. If I am to see a similarity between, say, my laughing or screaming and the laughing or screaming of somebody else, I need to understand the bodily gestures and behaviour as expressive phenomena, as manifestations of joy or pain, and not simply as physical movements. If such an understanding is required for the argument of analogy to proceed, however, the argument presupposes that which it is supposed to establish. In other words, we only employ analogical lines of reasoning when we are already convinced that we are facing minded creatures but are simply unsure about precisely how we are to interpret the expressive phenomena in question (Scheler 1973, p. 234; cf. Gurwitsch 1979, pp. 14, 18). In addition, Scheler also questions two of the crucial presuppositions that are made by the argument. First, the argument assumes that my point of departure is my own consciousness. This is what is at first given to me in a quite direct and unmediated fashion, and it is this purely mental self-experience that is then taken to precede and make possible the recognition of others. One is at home in oneself and one then has to project into the other, who one does not know, what one already finds in oneself. Second, the argument also assumes that we never have direct access to another person's mind. We can never *experience* her thoughts or feelings. We can only infer that they must exist based on that which is actually given to us, namely her bodily appearance. Although both of these two assumptions might seem perfectly obvious, Scheler rejects both. As he puts it, the argument from analogy underestimates the difficulties involved in self-experience and overestimates the difficulties involved in the experience of others (Scheler 1973, pp. 244–246). We should not ignore what can be directly perceived about others and we should not fail to acknowledge the embodied and embedded nature of self-experience. Thus, Scheler denies that our initial self-acquaintance is of a purely mental nature and that it takes place in isolation from others. But he also denies that our basic acquaintance with others is inferential in nature. Thus, as Scheler argues, there is something highly problematic about claiming that intersubjective understanding is a two-stage process of which the first stage is the perception of meaningless behaviour and the second an intellectually based attribution of psychological meaning. In the majority of cases, it is quite hard (and artificial) to divide a phenomenon neatly into its psychological and behavioural aspect, think merely of a groan of pain, a handshake, an embrace, a leisurely stroll. Scheler argues that we in the face-to-face encounter are neither confronted with a mere body, nor with a hidden psyche, but with a unified whole. He speaks of an "expressive unity" (*Ausdruckseinheit*). It is only subsequently, through a process of

abstraction, that this unity can be divided and our interest then proceeds "inwards" or "outwards" (Scheler 1973, p. 255).

Scheler opposes the view according to which our encounter with others is first and foremost an encounter with bodily and behavioural exteriorities devoid of any psychological properties. According to such a view, which has been defended by behaviourists and Cartesians alike, behaviour, considered in itself, is neither expressive nor significant. All that is *given* is physical qualities and their changes. Seeing a radiant face means seeing certain characteristic distortions of the facial muscles. But naturally, such a setup gives rise to the following sceptical question: How can we pass from a perception of the other as a "bag of skin moving over ground" (Gopnik and Meltzoff 1994, p. 166) to a full-blooded experience of the other as a minded creature?

However, as Scheler points out, this account presents us with a distorted picture, not only of behaviour but also of the mind. It is no coincidence that we use psychological terms to describe behaviour and that we would be hard pressed to describe the latter in terms of bare movements. In his view, affective and emotional states are not simply qualities of subjective experience, rather they are given *in* expressive phenomena, i.e. they are expressed in bodily gestures and actions, and they thereby become visible to others. This is, of course, not an idea to be found in Scheler alone:

> We must reject the prejudice which makes 'inner realities' out of love, hate or anger, leaving them accessible to one single witness: the person who feels them. Anger, shame, hate and love are not psychic facts hidden at the bottom of another's consciousness: they are types of behavior or styles of conduct which are visible from the outside. They exist *on* this face or *in* those gestures, not hidden behind them (Merleau-Ponty 1996, p. 67 [1964, pp. 52–53]).
>
> We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features (Wittgenstein 1980, Sect. 570).

Both Merleau-Ponty and Wittgenstein question the sceptic's conception of what is given. In Wittgenstein's view, the sceptics have displaced the human being with a philosophically generated concept of a human body, understood as a merely material object. Rather than attempting to solve the sceptical problem by somehow adding psychological meaning onto this impoverished object, he suggests that we instead simply restore the concept of a human being to its proper place, namely as a seamless whole of whose unity we should not have lost sight (cf. McDowell 1998, p. 384). Scheler would concur. Instead of attempting to secure an access to the minded life of others through technical detours, we need a new understanding of the given (cf. Gurwitsch 1979, pp. 29–30). If the realm of expressive phenomena is accepted as the primary datum or primitive stratum of perception, the access to the mind of others will no longer present the same kind of problem. To quote Scheler:

> For we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of his thoughts in the sound of his words. If anyone tells me that this is not 'perception', for it cannot be so, in view of the fact that a perception is simply a

'complex of physical sensations', and that there is certainly no sensation of another person's mind nor any stimulus from such a source, I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts (Scheler 1973, p. 254 [1954, p. 260]; cf. Gurwitsch 1979, p. 56).

It should by now be clear that Scheler takes a solution to the problem of other minds to require a correct understanding of the relation between mind and body. And of course, the mind-body relation in question is not the mind-brain relation. Scheler is not concerned with the search for the neural correlates of consciousness. Rather he is interested in the relation between experience and expressive behaviour.

2.3. EXPERIENCE AND EXPRESSION

Despite their differences, the theory theory of mind and the simulation theory of mind both deny that it is possible to *experience* other minded creatures; this is why we need to rely on and employ either theoretical inferences or internal simulations. Both accounts consequently share the view that the minds of others are hidden, and they consider one of the main challenges facing a theory of social cognition to be the question of how and why we start ascribing such hidden mental entities or processes to certain publicly observable bodies. As we have seen, phenomenologists would question this very setup. They would argue that the appeal to either theory or simulation is unwarranted since it is motivated by a too impoverished conception of what is given, of what is experientially available. It is occasionally assumed that a phenomenological account of intersubjectivity is by and large opposed to the theory theory of mind, whereas the relation between phenomenology and the simulation theory is much more conciliatory. I think this is only partially correct. In fact, by emphasizing the embodied nature of self-experience, several of the phenomenologists have called attention to a problem that in retrospect must seem particularly troublesome for Goldman's simulationist account. Whereas theory theorists such as Gopnik have traditionally emphasized the parallelism between the ascription of mental states to self and other (Gopnik 1993), Goldman has stressed the asymmetry. But if we start out by accepting the conceptual separation of the mental from the behavioural, and if my own self-experience is of a purely mental nature, whereas my experience of others is purely behavioural in nature, we need to understand why I should even so much as think that there are other minded creatures. As Davidson has put it:

If the mental states of others are known only through their behavioral and other outward manifestation, while this is not true of our own mental states, why should we think our own mental states are anything like those of others? (Davidson 2001, p. 207).

If we adopt what McCulloch has recently called a behaviour-rejecting mentalism (McCulloch 2003, p. 94), i.e. if we deny that embodiment and bodily behaviour have any essential role to play in experience and cognition, if we deny that embodiment and environmental embedding are essential to having a mind, we will have a hard time escaping what is known as the *conceptual problem of other minds* (cf. Avramides 2001).

The proper way to respond to this sceptical challenge is by abandoning the radical divide between the subject's mind and body. This is where the notions of expression and action become crucial. It could be argued, of course, that any account of the mind has to take subjectivity and the first-person perspective seriously, and that a focus on expression and action will consequently lose sight of what is essential to the mind. However, this worry is misguided. There is nothing reductive in the reference to expression and action, since subjectivity figures centrally in both concepts.

The idea is not to reduce consciousness as such to intentional behaviour. But we should recognize that the expressive relation between mental phenomena and behaviour is stronger than that of a mere contingent causal connection, though weaker than that of identity. The behaviour is neither necessary nor sufficient for a whole range of mental phenomena, so one can occur without the other – which is why deception and suppression are possible – but this is not to say that this is generally the case or that it could conceivably always be the case. As a rule, we do not come to know one independently of the other. In fact, as Rudd has recently argued, intersubjective understanding is possible precisely because some of our mental states find a natural expression in bodily behaviour, and because the language we learn for our mental states is a language that we learn to apply to others even as we learn to apply it to ourselves (Rudd 2003, p. 114).

Our understanding of affective and emotional states, such as sorrow, shame, love, gratitude, hate, pity, disgust, fear, pride, etc., is informed and influenced by their behavioural manifestations. The latter is partly constitutive of the meaning of the former. As Cassirer formulated it in his *Philosophie der symbolischen Formen*, "Life cannot apprehend itself by *remaining* absolutely within itself. It must give itself form; for it is precisely by this 'otherness' of form that it gains its 'visibility,' if not its reality" (Cassirer 1954, p. 46 [1957, p. 39]. Cf. Cassirer 1954, pp. 43–47).

Expression is more than simply a bridge supposed to close the gap between inner mental states and external bodily behaviour. Some mental states are directly apprehended in the bodily expressions of people whose mental states they are, or as Hobson has recently put it: "We perceive bodies and bodily expressions, but we do so in such a way that we perceive and react to the mental life that those physical forms express" (Hobson 2002, p. 248). In seeing the actions and expressive movements of other persons, one already sees their meaning. No inference to a hidden set of mental states is necessary. Expressive behaviour is soaked with the meaning of the mind; it reveals the mind to us. Certainly, it differs from the direct manifestation available from the first-person perspective. We should respect and maintain the asymmetry between the first-person and the second- (and third-) person access to psychological states, but this is not a difference between an immediate certainty on the one side and an insecure inference on the other. As Wittgenstein writes, "My thoughts are not hidden from [the other], but are just open to him in a different *way* than they are to me" (Wittgenstein 1992, pp. 34–35). Nor should we make the mistake of confusing different kinds of access with different degrees of certainty. As Wittgenstein also points out, even if I had no uncertainty with regard to the mental state of an other (say, in the case where I observe the victim of a

car accident writhing in pain), that would not make it *my* state (Wittgenstein 1982, Sect. 963). We should recognize that each type of access has its own strengths and weaknesses. The second- (or third-) person access only "falls short" of the first-person access, if it is assumed that the latter is privileged, and that it is the internal aspiration of the former to approximate the latter as closely as possible (Moran 2001, p. 157).

We should oppose the view that behaviour, considered in itself, is neither expressive nor significant. We should oppose the view that behaviour is merely the outwardly observable effects of mental states and goings-on. Such a view does not merely fail to recognize the true nature of behaviour, it also presents us with a misleading perspective on the mind, suggesting as it does that the mind is a purely internal happening located and hidden in the head, thereby giving rise to the sceptical conundrum (cf. McDowell 1998, p. 393). We should avoid construing the mind as something visible to only one person and invisible to everyone else. The mind is not something exclusively inner, something cut off from the body and the surrounding world. As if psychological phenomena would remain exactly the same, even without gestures, bodily expressions, etc. As Overgaard points out, psychological phenomena stretch their arms in many directions – they play many publicly observable roles – and to cut off all these public arms would leave us with a severely distorted picture of the mental (Overgaard 2005).

Our ascription of conscious states to others is based on behavioural evidence. But this is not to say that the ascription is hypothetical or assumptive and our understanding indirect or inferential. As Bennett and Hacker observe, when somebody blushes because he is ashamed, the blush reveals and manifests the shame; it doesn't conceal it. When somebody screams in pain while the dentist drills in his tooth, it makes little sense to say that this is merely behaviour and that the real pain is still concealed and inner. We can speak of indirect evidence or of knowing indirectly only where it also makes sense to speak of a more direct evidence, but there is no more direct way of knowing that somebody is in pain than seeing him writhe in pain, that somebody sees something than by him showing what he sees and knowing what he thinks than from his sincere confession. Knowing indirectly or by way of inference that somebody is in pain might be a matter of noticing a bottle of pain killers next to his bedside together with an empty glass of water (Bennett and Hacker 2003, pp. 89, 93).

To repeat, this is not behaviourism. The idea is not to identify mental states with or reduce them to behaviour, nor does it rule out that some experiential states are covert, but not all experiences can lack a natural expression if intersubjectivity is to get off the ground.[3] To suggest that the indirect means of verifying claims about black holes or subatomic particles can "give us a model for verifying hypotheses in the area of the study of human and animal subjectivity" (Searle 1999, p. 2074) is deeply confused.

Whereas we in adult life occasionally make inferential attributions of mental states to other people, such attributions cannot be considered the basis of the smooth and immediate interpersonal interaction found in infants (Trevarthen 1979). A vivid

example of the importance of facial expressions is provided by the famous "visual cliff" experiment. Infants aged 12 months are placed on one side of a "visual cliff", i.e. an apparent sudden drop beneath a transparent surface. On the other side of the cliff, the infant's mother and an attractive toy are placed. When the infant notices the drop-off, it will typically look spontaneously at the mother's face. If the mother poses a happy face, most infants will cross to the other side; if the mother poses a fearful expression, the infants will freeze or even actively retreat. As Hobson points out, it is noteworthy that the mother's mere presence is not enough, rather her emotional reaction, as perceived through her expressions and behaviour, has a decisive influence (Hobson 1991, p. 47). In other words, the infant appears to recognize that another person's expression has meaning with reference to an environment common to both of them. The infant is not living in a solipsistic world, a world that has the meaning it has solely because of how it is taken by the infant. Rather, the world has also meaning for others, and the meaning it has for others affects the meaning it has for the infant. Thus, the gestures and utterances of the caretaker are perceived as being both emotionally expressive and as being directed to something in the infant's world (Hobson 1993, pp. 38, 140–141, 2002, p. 73). This makes Hobson conclude that infants "have direct perception of and natural engagement with person-related meanings that are apprehended in the expressions and behaviour of other persons. It is only gradually, and with considerable input from adults, that they eventually come to conceive of 'bodies' on the one hand, and 'minds' on the other" (Hobson 1993, p. 117).

Tomasello has recently proposed that our social cognition takes three forms. We can understand others (1) as animate beings, (2) as intentional agents, and (3) as mental agents. In his view, the ontogenetic relevance of this tripartition is straightforward. Whereas infants are able to distinguish animate beings from non-animate beings already from birth onwards, they are only able to detect intentionality, in the sense of goal-directed behaviour, from around 9–12 months of age (as evinced from phenomena such as joint attention, gaze following, joint engagement, imitative learning, etc.), and they only become aware of others as mental agents with beliefs that might differ from their own at around 4–5 years of age. Why does the last step take so much longer? The answer provided by Tomasello is twofold. On the one hand, he calls attention to the different roles of expressive behaviour. Whereas the animacy of others is directly expressed in their behaviour, intentionality is also expressed in actions, but is at the same time somewhat divorced from them, since it on occasions may remain unexpressed or be expressed in different ways. Finally when it comes to thoughts and beliefs, these might lack natural behavioural expressions altogether (Tomasello 1999, p. 179), which is what makes them so much more difficult to grasp. On the other hand, Tomasello argues that the more advanced form of social cognition emerges as late as it does because it depends on prolonged real-life social interaction (Tomasello 1999, p. 198). More specifically, he argues that language use may play a crucial role in children's coming to view other people as mental agents (Tomasello 1999, p. 176). In order to understand that other persons have beliefs about the world that differ from their own, children

need to engage them in discourses in which these different perspectives are clearly apparent, be it in disagreements, misunderstandings, requests for clarification or reflective dialogues (Tomasello 1999, pp. 176, 182). Although I find it potentially misleading to designate the difference between an understanding of the goal-directed actions of others and an understanding of their false or divergent beliefs as a difference between an understanding of others as intentional agents and as mental agents – it might suggest both that there is nothing mindful about goal-directed actions and no intentionality to thoughts and beliefs – Tomasello is certainly right in pointing to the fact that our understanding of others gradually becomes more sophisticated and that there are dimensions of the mind that are not as readily accessible as others. Moreover, I also think he is right in pointing to the cultural and social dimension of this developmental process. Rather than being the result of an automatic maturation of certain innate cognitive modules, it seems plausible to view these more sophisticated forms of social cognition as something that develops in tandem with increasingly complex forms of social interaction.

2.4. THE ROLE OF EMPATHY

One reason why the problem of other minds seems so persisting is that we have conflicting intuitions about the accessibility of the mental life of others. On the one hand, there is something right about the claim that the feelings and thoughts of others are manifest in their expressions and gestures. On the other hand, there also seems to be something right in the idea that the mental life of another is in some respect inaccessible. There are situations where we have no reason to doubt that the other is angry, in pain, or bored. There are other situations where we have no clue as to their precise state of mind. It seems wrong to claim that the mental life of others is essentially inaccessible, but it also seems wrong to claim that everything is open to view. The challenge is to reconcile both intuitions, rather than letting one of them go (Overgaard 2005).

Most phenomenologists have argued that it makes no sense to speak of another unless the other is in some way given and accessible. That I have an actual experience of the other, and do not have to do with a mere inference or imaginative simulation, does not imply, however, that I can experience the other in the same way as she herself does, nor that the other's consciousness is accessible to me in the same way as my own is. The second- (and third-) person access to psychological states differs from the first-person access, but this difference is not an imperfection or a shortcoming. Rather, the difference is constitutional. It is what makes the experience in question an experience of another, rather than a self-experience. As Husserl writes, "Had I had the same access to the consciousness of the other as I have to my own, the other would have ceased being an other, and instead have become a part of myself" (Husserl 1950, p. 139). To put it differently, the first-personal givenness of the mind of the other is inaccessible to me, but it is exactly this inaccessibility, this limit, which I can experience, and which makes the experience in question an experience of another (Husserl 1950, p. 144). We experience the

behaviour of others as expressive of mental states that *transcend* the behaviour that expresses them. Thus, the givenness of the other is of a most peculiar kind. The otherness of the other is exactly *manifest* in his elusiveness and inaccessibility. As Lévinas observes, "the absence of the other is exactly his presence as other" (Lévinas 1979, p. 89). To demand more, to claim that I would only have a real experience of the other if I experienced her feelings or thoughts in the same way as she herself does, is nonsensical. It would imply that I would only experience an other if I experienced her in the same way that I experience myself, i.e. it would lead to an abolition of the difference between self and other, to a negation of the *alterity* of the other, of that which makes the other.

What needs to be realized is that bodies of others differ radically from inanimate objects, and that our perception of these minded bodies is unlike our ordinary perception of objects. The relation between self and other is not first established by way of an inference to the best explanation, nor is it a question of putting ourselves in the shoes of the other. On the contrary, we should recognize the existence of a distinctive mode of consciousness, different from perception, recollection and fantasy, that permits us to understand others. The traditional name for this mode of experience is *empathy*.

The use of the concept of empathy is not uncontroversial. In fact, in phenomenological circles, the notion has fallen into a certain disrepute because of Heidegger's well known criticism. As Heidegger argues, if one seeks to understand intersubjectivity on the basis of empathy one will remain committed to a serious misconception of the nature of the self:

> If this word [empathy] is at all to retain a signification, then only because of the assumption that the "I" is at first in its ego-sphere and must then subsequently enter the sphere of another. The "I" does not first break out […] since it already is outside, nor does it break into the other, since it already encounters the other outside (Heidegger 2001, p. 145).

According to this understanding of the concept, the notion of empathy is linked to the problem of how one (isolated) subject can encounter and understand another (isolated) subject. Even if the empathic approach does not commit the same mistakes as the argument from analogy, it still misconstrues the nature of intersubjectivity since it takes it to be first and foremost a thematic encounter between individuals, where one is trying to grasp the emotions or experiences of the other (this connotation is particularly obvious in the German word for empathy: *Einfühlung*). However, as Heidegger also points out, the very attempt thematically to grasp the experiences of others is the exception rather than the rule. Under normal circumstances, we understand each other well enough through our shared engagement in the common world.

A similar criticism can be found in Gurwitsch. Gurwitsch readily acknowledges the importance of expressive phenomena, but he criticizes Scheler for having been too one-sided in his approach and then argues that the realm of expressive phenomena is neither the only, nor the primary, dimension to be considered if we wish to understand what it is that enables us to encounter other human beings as humans (Gurwitsch 1979, p. 33).

Ordinarily, we do not encounter others primarily as thematic objects of cognition. Rather we encounter them in the world in which our daily life occurs, or to be more precise, we encounter others in a worldly situation, and our way of being together and understanding each other is co-determined in its meaning by the situation at hand (Gurwitsch 1979, pp. 35–36, 95, 106). To exemplify, Gurwitsch analyses the situation where two workers are cobbling a street. In this work situation, one worker lays the stones while the other knocks them into place. Each worker is related to the other in his activity and comportment. When one worker understands the other, the understanding in question does not involve grasping some hidden mental occurrences. There is no problem of other minds. There is no problem of how one isolated ego gets access to another isolated ego. Rather both workers understand each other in virtue of the roles they play in the common situation (Gurwitsch 1979, pp. 104, 108, 112).

It is precisely within such common situations that expressive phenomena occur. When working or conversing with my partner, he might shake his head or wrinkle his brow. But these facial expressions and bodily gestures are not unambiguous. They do not reveal psychological states simply or uniformly. Each person has different countenances and facial habits. But this is rarely a problem, since we do not encounter expressions in isolation. They always occur in a given context, and our understanding of the context, of what comes before and after, helps us understand the expression. As Gurwitsch points out, the "same" shaking of the head can take on different meanings in different situations. What an expressive phenomenon is and signifies in a particular case becomes comprehensible to me in the whole of the present situation (Gurwitsch 1979, p. 114; cf. Sartre 1943/1976, p. 396).

Heidegger and Gurwitsch would both deny that intersubjective understanding is primarily or even ordinarily a question of successfully ascribing causally efficacious inner mental states (mainly beliefs and desires) to others in order to facilitate our prediction and explanation of their behaviour. They would concur with Hutto's recent claim that explanation and prediction of action from a third-person stance is far more infrequent and far less reliable than our normal intersubjective means of coming to understand others through dialogue and conversation and shared narratives (Hutto 2004). If somebody is acting in a puzzling way, by far the easiest and most reliable way to gain further information is not to engage in detached theorizing or internal simulation, it is to employ your conversational skills and ask the person for an explanation.

However, I think one can accept these critical points and still consider the notion of empathy to be useful. One should simply acknowledge that our typical understanding of others is contextual and realize that empathy, properly understood, is not a question of feelingly projecting oneself into the other, but rather an ability to experience behaviour as expressive of mind, i.e. an ability to access the life of the mind of others in their expressive behaviour and meaningful action.

We should respect the difference between the first-person and the second- and third-person perspectives and we should recognize the difference between self-

experience and the experience of others. But too much focus on this difference or asymmetry can lead to the mistaken view that only my own experiences are given to me and that the behaviour of the other shields his experiences from me and makes their very existence hypothetical (Avramides 2001, p. 187).

Our experience and understanding of others are not infallible. Other people can fake or conceal their experiences. But there is a decisive difference between our everyday uncertainty about what precisely others might be thinking about and the nightmare vision of the solipsist. Although we might be uncertain about the specific beliefs and intentions of others, this uncertainty does not make us question their very existence. In fact, as Merleau-Ponty points out, our relation to others is deeper than any specific uncertainty we might have regarding them (Merleau-Ponty 1945, p. 415).[4]

2.5. CONCLUSION

In contrast to the take favoured by simulationists and theory theorists alike, the crucial question is not whether we can predict and explain the behaviour of others, and if so, how that happens, but rather whether such prediction and explanation constitute the primary and ordinary form of intersubjectivity. There is a marked difference between the way we engage with others in the second-person and the third-person case. When we interact directly with another person, we do generally not engage in some detached observation of what the person is doing. We do in general not at first attempt to classify his or her actions under lawlike generalizations; rather we seek to make sense of them. When you see somebody use a hammer, feed a child or clean a table, you might not necessarily understand every aspect of the action, but it is immediately given as a meaningful action (in a common world). Under normal circumstances, we understand each other well enough through our shared engagement in this common world, and it is only if this pragmatic understanding for some reason breaks down, for instance if the other behaves in an unexpected and puzzling way, that other options kick in and take over, be it inferential reasoning or some kind of simulation. We develop both capacities, but we only employ them in special circumstances. Neither establishes our primary nor ordinary access to the embodied minds of others. They are the exceptions rather than the rules. In most intersubjective situations, we have a direct understanding of the other person's intentions, since these intentions are manifested in the person's behaviour and embedded in a shared social context. Thus, as Gallagher remarks, much is going on in our understanding of others that exceeds and precedes our theoretical and simulation capabilities. At best, the theory theory of mind and the simulation theory of mind only explain a narrow and specialized set of cognitive processes that we can employ when our usual ways of understanding others fall short (Gallagher 2005, p. 208).

The simulation theory of mind and the theory theory of mind both operate with a problematic dichotomy between inner and outer, between experience and behaviour. But if we start with a radical division between a perceived body and an inferred

or simulated mind, we will never, to use Hobson's phrase, be able to "put Humpty Dumpty together again" (Hobson 1993, p. 104).

ACKNOWLEDGEMENTS

Thanks to Dan Hutto and Matthew Ratcliffe for some helpful comments. This study was funded by the Danish National Research Foundation.

NOTES

[1] Let me add though that I actually have considerable more sympathy for some of the other "simulationist" approaches. I think that Gordon's reference to ascent routines and Gallese's investigations of the automatic, implicit and subpersonal activation and resonance of mirror neurons both contain important insights. It is simply less clear to me what either has to do with simulation in the proper sense of the word. Thus I would, for instance, agree with Gallagher when he, elsewhere in this volume, argues that mirroring rather than being a question of simulation is a question of perceptual elicitation, and when he writes that it therefore makes better sense to consider it as a process that subserves our direct understanding of the actions of others (Gallagher 2007).

[2] In Zahavi (2004, 2005), I discuss this question in more detail.

[3] It might be worth mentioning that some empirical research suggests that the expression of a number of basic emotions, such as anger, happiness, disgust, contempt, sadness, fear and surprise, are cross cultural and universal, though there are, of course, culturally specific rules about how to manage expressions in public (Ekman 2003, pp. 4, 10, 58). The suggestion that basic emotional expressions are innate is further corroborated by the fact that even congenitally blind children normally exhibit the relevant facial expressions.

[4] For a more extensive discussion of phenomenological accounts of intersubjectivity, cf. Zahavi (2001a, b).

REFERENCES

Avramides A (2001) Other minds. Routledge, London.

Baron-Cohen S (1995) Mindblindness. An essay on autism and theory of mind. MIT Press, Cambridge, MA.

Bennett MR, Hacker PMS (2003) Philosophical foundations of neuroscience. Blackwell, Oxford.

Cassirer E (1954) Philosophie der symbolischen Formen III. Wissenschaftliche Buchgesellschaft, Darmstadt; The philosophy of symbolic forms III. Transl. by Manheim R (1957) Yale University Press, New Haven.

Davidson D (2001) Subjective, intersubjective, objective. Oxford University Press, Oxford.

Ekman P (2003) Emotions revealed: understanding faces and feelings. Weidenfeld & Nicolson, London.

Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford.

Gallagher S (2007) Logical and phenomenological arguments against simulation theory. In: Hutto D, Ratcliffe M (eds) Folk psychology reassessed. Springer, Berlin Heidelberg New York.

Goldman AI (2000) Folk psychology and mental concepts. Protosociology 14:4–25.

Gopnik A (1993) How we know our minds: the illusion of first-person knowledge of intentionality. Behav Brain Sci 16:1–14.

Gopnik A, Meltzoff AN (1994) Minds, bodies and persons: young children's understanding of the self and others as reflected in imitation and 'theory of mind' research. In: Parker S, Mitchell R (eds) Self-awareness in animals and humans. Cambridge University Press, New York, pp 166–186.

Gordon RM (1996) 'Radical' simulationism. In: Carruthers P, Smith PK (eds) Theories of theories of mind. Cambridge University Press, Cambridge, pp 11–21.

Gurwitsch A (1979) Human encounters in the social world. Transl. by Kersten F. Duquesne University Press, Pittsburgh, PA.
Heidegger M (2001) Einleitung in die Philosophie. Frankfurt am Main. Vittorio Klostermann.
Hobson RP (1991) Against the theory of 'Theory of Mind'. Br J Dev Psychol 9:33–51.
Hobson RP (1993) Autism and the development of mind. Psychology Press, Hove.
Hobson RP (2002) The cradle of thought. Macmillan, London.
Husserl E (1950) Cartesianische Meditationen und Pariser Vorträge. Martinus Nijhoff, Den Haag.
Hutto DD (2004) The limits of spectatorial folk psychology. Mind Lang 19/5:548–573.
Leslie AM (1987) Children's understanding of the mental world. In: Gregory RL (ed) The Oxford companion to the mind. Oxford University Press, Oxford, pp 139–142.
Lévinas E (1979) Le temps et l'autre. Fata Morgana, Paris.
McCulloch G (2003) The life of the mind: an essay on phenomenological externalism. Routledge, London.
McDowell J (1998) Meaning, knowledge, and reality. Harvard University Press, Cambridge, MA.
Merleau-Ponty M (1945) Phénoménologie de la perception. Éditions Gallimard, Paris.
Merleau-Ponty M (1996) Sens et non-sens. Èditions Gallimard, Paris; Transl. by Dreyfus HL, Dreyfus PA (1964). Northwestern University Press, Evanston, IL.
Moran R (2001) Authority and estrangement: an essay on self-knowledge. Princeton University Press, Princeton, NJ.
Overgaard S (2005) Rethinking other minds: Wittgenstein and Levinas on expression. Inquiry 48/3: 249–274.
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 4:515–526.
Rudd A (2003) Expressing the world: skepticism, Wittgenstein, and Heidegger. Open Court, Chicago, IL.
Sartre J-P (1943/1976) L'être et le néant. Tel Gallimard, Paris.
Scheler M (1973) Wesen und Formen der Sympathie. Francke Verlag, Bern; The nature of sympathy. Transl. by Heath P (1954). Routledge & Kegan Paul, London.
Searle JR (1999) The future of philosophy. Philos Trans R Soc Lond B 354:2069–2080.
Tomasello M (1999) The cultural origins of human cognition. Harvard University Press, Cambridge, MA.
Trevarthen C (1979) Communication and cooperation in early infancy: a description of primary intersubjectivity. In: Bullowa M (ed) Before speech. The beginning of interpersonal communication. Cambridge University Press, Cambridge, pp 321–347.
Wellman HM, Cross D, Watson J (2001) Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev 72:655–684.
Wittgenstein L (1980) Remarks on the philosophy of psychology II. Blackwell, Oxford.
Wittgenstein L (1981) Zettel. Blackwell, Oxford.
Wittgenstein L (1982) Last writings on the philosophy of psychology, vol. 1. Blackwell, Oxford.
Wittgenstein L (1992) Last writings on the philosophy of psychology, vol. 2. Blackwell, Oxford.
Zahavi D (2001a) Beyond empathy. Phenomenological approaches to intersubjectivity. J Conscious Stud 8(5–7):151–167.
Zahavi D (2001b) Husserl and transcendental intersubjectivity. Ohio University Press, Athens.
Zahavi D (2004) The embodied self-awareness of the infant: a challenge to the theory theory of mind? In: Zahavi D, Grünbaum T, Parnas J (eds) The structure and development of self-consciousness: interdisciplinary perspectives. Benjamins, Amsterdam, pp 35–63.
Zahavi D (2005) Subjectivity and selfhood: investigating the first-person perspective. The MIT Press, Cambridge, MA.

3. WE SHARE, THEREFORE WE THINK

3.1. INTRODUCTION

Elsewhere in this volume one can find telling critiques of contemporary accounts of the nature and scope of interpersonal understanding. Therefore, apart from a few unsubtle but I believe important introductory points to provide some initial orientation to my own approach, I shall not attempt to present or analyze views about human beings' knowledge of other people with which I am in disagreement. Rather, I shall consider some things about our ways of knowing others, and more specifically, our ways of coming to know about other people with minds, that might re-frame and perhaps restructure current theories in this domain.

Here is the direction from which I shall be coming. A principal concern of my research as an experimental psychologist has been to investigate the nature of impairments in interpersonal relations, communication and self-awareness among children and adolescents with autism. Colleagues and myself have also studied related issues among congenitally blind children, as well as aspects of social functioning among typically developing infants (sometimes infants of troubled mothers). So it seems fitting that I should draw upon some findings from this research programme, insofar as they may highlight what might otherwise be neglected in abstract theorizing about interpersonal relations and understanding.

Of course empirical investigations are no substitute for conceptual analysis of the things we need to explain. Yet one motive that gives impetus to our research is an interest in genetic epistemology (Piaget 1972) as this applies to a particular topic, namely, the developmental conditions that make the acquisition of knowledge about human minds possible. More specifically, we have been concerned that attempts not only to explain the syndrome of autism, but also to account for early stages of cognitive development in typically developing young children, are often distorted by a failure to appreciate the formative influence of mutual, affectively patterned interpersonal relations on human psychological growth.

3.1.1. A perspective on interpersonal understanding

It has long perplexed me that psychologists and philosophers have been happy to adopt and even espouse the phrase 'theory of mind' to characterize human interpersonal understanding. I wrote a paper on this a few years ago (Hobson 1991), where among the arguments I rehearsed was that a young child would hardly be able to assume the stance of the little scientist and derive hypotheses about the mind and its workings unless he or she were in a position to frame those hypotheses in the first place. The abilities to articulate and entertain alternative and hypothetical

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 41–61.

descriptions of states of affairs, never mind test them out, is central to what we need to explain. A move that by-passes this concern is to argue that theory of mind is innate, but one among several difficulties here is to justify the claim that whatever the mind's inbuilt propensities for interpreting psychological relations, these are derived not from perceptual sensibilities but from the deployment of theory-like, cognitively structured modes of understanding (also Gallagher, this volume). Of course one might object that the use of the term 'theory' was never meant to be literal, but merely served to highlight certain features of knowledge that stand in contrast to alternative theoretical perspectives on how we come to think about mental states (e.g. Gopnik and Meltzoff 1997). In this case, it may be worth considering whether what we gain from such an analogy is substantial enough to justify its use, when it so threatens to skew our perspective.

In my view, there is one great plus to such a stance and lots of minuses. The plus is the insistence that we need to give a developmental account of the acquisition of *concepts* concerning the mind. The minuses include all those things implicated in a conception of the little scientist as a more or less isolated individual set against the social world, trying to work out and explain things to himself or herself. Instead, one might see infants and toddlers as fully human beings with intense and feeling-full *engagements with* others and the world, young children whose understanding and conceptual abilities in general, as well as whose understanding and concepts about people, need to be grounded in something beyond the kinds of intellectual operation encompassed in the notion of 'theorizing', however elastically employed. Indeed, when it comes to explaining the patterning of *infants'* relations with other people, developmentally inclined theory theorists have tended to characterize the earliest modes of understanding by reference to intuitive theory of mind abilities, or precursors to theory of mind, a strategy that maintains theory as the fulcrum of development. In addition, aspects of social perception or sensitivity such as emotion recognition have come to be subsumed under the same theory of mind rubric. So it is that substantive and helpful theoretical distinctions that feature in theory of mind theorizing are being smudged out of recognition.

There has been another curious side effect of all this, namely the polarization into a battle between two camps, the theory of mind theorizers and the simulationists. Crudely expressed (and see Gallagher, this volume, for a subtle and penetrating analysis) simulation is supposed to be the means by which one person puts himself into the shoes of another, and thereby achieves understanding of the other according to the model of himself. Not a few simulationists seem to have been ensnared by the 'argument from analogy' demolished (in my view) by Wittgenstein (1958), as they have imagined that one might somehow arrive at concepts of mental states concerning one's own case prior to, and as a precondition of, ascribing them to others. If one arrives at knowledge that someone else has such-and-such a belief or desire by imagining oneself in their position, then one needs to be able not just to imagine (a significant developmental achievement that requires interpersonal understanding, I shall argue), but moreover to imagine by drawing upon concepts of one's own mental states. Here the dangers have been to underestimate the

importance of social-communicative factors in accounting for the acquisition of concepts of mind, as well as to proffer such a pale information processing type account of simulation as to gainsay the qualities and developmentally formative nature of interpersonal engagement.

As I see it, then, we might do better to distil what is valuable in each of these two views and start again. There is indeed a challenge to understand the nature and development of concepts of mind, and there are indeed modes and levels of interpersonal connection and responsiveness that do not entail theorizing. More or less self-evidently, there are also ways of understanding others that involve reasoning by 'putting oneself in another person's shoes', but only once one has the imaginative and conceptual equipment to reason in this way.

So the question arises: How might an alternative perspective on the growth of interpersonal understanding be fashioned?

We should start by taking seriously the qualities of relatedness that can exist between one person and another. We love and hate each other, we are drawn to engage with or are repelled by one another, we compete and cooperate, we share things or reject the opportunities to share, and although we envy and we covet, we also give things to others, show concern, and so on (also Andrews, this volume). Broadly speaking, people also have much in common when it comes to relating to the non-personal world. Not surprising, then, that we come to understand how we – ourselves and others – relate to each other and to our shared world in just these typically human ways. But although it seems self-evident that all this is relevant for our understanding of how people tick, is it so important for our understanding of *minds*, or perhaps more specifically, for the concepts in terms of which our understanding of minds is framed?

This is a curious and misleading question, in that it creates a split between understanding persons and understanding minds. But now that I have posed the question, I had better respond to it. I shall do so by giving three reasons, each very different from the others, for believing that such relations should figure prominently in our account of how children acquire mental concepts.

The first reason is that what one observes in human development from infancy onwards, and what one finds specifically affected among children with autism (the relevance of which will become clear), is the kind of distinction between I-Thou and I-It relations that is especially apparent in the patterns of affectively coordinated exchange between a human being and others (see also Stawarska, this volume). One way of making the point would be to say that even in infancy, children discriminate between people and things in their qualities of relatedness. As we shall see, these qualities become more complex as development proceeds. Why not suppose that what *look like* foundations for what will become more articulated understandings of the nature of people vis-à-vis things, are indeed what they seem?

By way of disagreement, some might say that early interpersonal relations, or the qualities of relatedness that are already apparent in infancy and that continue through life, might have little bearing on the later emergence of concepts of mind (also Hutto 2004). There might be independent cognitive/computational mechanisms

that are primarily responsible for the so-called theory of mind, so that earlier appearing and supposedly more basic processes of social-affective relations are either beside the point or merely there to be 'triggered' into supporting higher-order representations (e.g. Leslie 1987). Even if one were to say that infants manifest an implicit 'understanding' of people at the end of the first year of life, for example in episodes of joint attention where they bring objects to share with others, this might occur through instinctively based patterns of interaction that are later replaced by or overlaid with truly representational abilities that have a quite different source. Yet if our aim is to understand how children come to conceptualize *people's* minds, rather than explain the emergence of mental representations of mental representations, then developmental continuity in different forms of interpersonal relations, only the later variants of which entail conceptual understanding, seems a much more attractive option. Indeed I shall argue that a young child's increasingly elaborated relations with, and increasingly articulated understanding of, bodily endowed people is foundational for acquiring concepts of mind. Infants begin with preconceptual modes of perceiving and relating to people, and on this basis come to acquire initial conceptual understanding of their own and others' minds in the second year of life.

My second reason has to do with what it *feels like* to relate to people, that is, to share with, to learn from, to compete, to trust, to envy, to love and hate, to need, to lose and to do and feel all kinds of other things with people. I have been convinced by the writings of Hamlyn (1974, 1978) in particular, that in order to have a concept of persons (with minds), it is necessary to experience and understand the kinds of relations that can exist between oneself and people, and especially, relations based on mutual feelings. A moment's reflection on one's own experiences in relation to others suggests that indeed, the ways others affect us and we affect them are somehow basic for our grasp of what other people *are* and how we understand what they are about.

Again it is possible that such feelings towards, with and from others might be thought to arise only as a result of developments in mental representation and cognitive growth. Or one might suppose that such features of interpersonal life may fill out our grasp of the contents of people's wishes, intentions, beliefs, and so on, but not provide the grounds for conceptualising such mental states, that is, not configure the intellectual means by which we think about the mind. It would certainly be bold and probably foolhardy to claim such attitudes are sufficient for deriving the necessary forms of conceptual insight, since there is much else in a child's transactions with the non-personal as well as personal world that feeds into the growth of understanding 'in general'. Yet I believe that when the *inter*personal transactions that implicate such feelings are taken into account, these provide much that is needed for growth in the intellectual structures as well as content of understanding persons with minds. In particular, it is through an infant's emotional engagement with other people's attitudes towards a shared world towards the end of the first year of life, that the infant is lifted out of his or her own one-track mode of apprehending objects and events; and it is through this mode of non-inferential, non-conceptual role-taking that over the coming months, the toddler

acquires the *concept* of what it is to be a self who entertains a person-anchored perspective. This amounts to a grasp of the distinction between thought and what is thought about, a distinction that is critical for the structuring of thought and symbolic representation (Hobson 1993a, 2000, 2002).

Perhaps it is already clear that, following Strawson (1962), I take it that fundamental to our thinking about minds is the concept of person, the kind of thing that has both a body and mind. What one observes in infants is their sensitivity to what is expressed *through* the bodily behaviour and expressions of other people, so that they apprehend subjective life through their feeling perception of others (also Zahavi, this volume). Only later, of course, do they conceive of mental states, and again such concepts of mind have essential links with concepts of embodied persons or selves whose minds they are. In terms of genetic epistemology, it is not so much a matter of hooking-up our mentalistic understanding with bodily goings-on, but rather, distilling out from relational events understandings of subjective and bodily dimensions of personal life. So I was amused when recently a colleague framed a topic for discussion by a study group as: 'How does the mind come to have the concept of self?' For me, the question is: 'How does a person come to acquire concepts, including concepts of the mind and the self?' (Hobson 1990). As so often in science and philosophy, the way in which our questions are framed determines the shape of the answers we offer. It is essential to our concepts of mind, that they are ascribable to persons or selves whose minds they are, and we might have less difficulty in fitting things together if we appreciate that it is only in our ways of thinking that bodies and minds have become partly (but only partly) separated (Hobson 1990, 1993b). Of course, if we need a self in order to represent, then it will be necessary to analyze the conditions that make it possible to have whatever kind of self accomplishes this.

3.1.2. The case of autism: one-to-one interpersonal engagement

Here I offer some clinical material and examples of studies of children and adolescents with autism, to highlight aspects of their interpersonal relations that might draw attention to what we are likely to overlook as important for the growth of social understanding. For the moment, I am going to assume it is the case that children with autism are limited in their understanding of minds. Most psychologists would take this to be epitomized by the children's difficulties in passing tests of understanding what it means to have false beliefs or knowledge (originally, Baron-Cohen et al. 1985). However constricted this view may be, it is helpful for the present purposes, in the sense that it motivates us to examine what might underpin such difficulties in understanding. I do accept that children with autism have limited knowledge in these respects, and that it is important they do so. So the following could be read as sketching a view of how concepts of false belief and knowledge are so difficult for children with autism to acquire, even though it is a view with far broader horizons.

I am going to give three brief snapshots of what children with autism are like in their one-to-one, dyadic interpersonal relations. The first is a clinical description

from the classic paper by Kanner (1943) in which the syndrome of autism was first described. Here is a description of one of Kanner's cases:

Case 9: Charles was brought to the clinic at the age of four and a half years, his mother complaining how "the thing that upsets me most is that I can't reach my baby". As a baby, this child would lie in the crib, just staring. When he was one and a half years old, he began to spend hours spinning toys and the lids of bottles and jars. His mother remarked: "He would pay no attention to me and show no recognition of me if I enter the room. The most impressive thing is his detachment and his inaccessibility. He walks as if he is in a shadow, lives in a world of his own where he cannot be reached. No sense of relationship to persons. He went through a period of quoting another person; never offers anything himself. His entire conversation is a replica of whatever has been said to him. He used to speak of himself in the second person, now he uses the third person at times; he would say, "He wants" – never "I want"… When he is with other people, he doesn't look up at them. Last July, we had a group of people. When Charles came in, it was just like a foal who'd been let out of an enclosure… He has a wonderful memory for words. Vocabulary is good, except for pronouns. He never initiates conversation, and conversation is limited, extensive only as far as objects go."

My second snapshot is of a quasi-observational study of adolescents with autism. In an attempt to capture something of what interpersonal engagement (or its relative lack) means, my colleague Tony Lee and I decided to study what happens in entrances to and exits from such social contact in the course of greetings and farewells (Hobson and Lee 1998). In what we call our Hello-Goodbye study, we videotaped a group of children and adolescents with autism and a group of matched non-autistic individuals in a standardized situation where they were introduced to, and subsequently departed from, a stranger seated across the room from the door by which they entered and left. Compared with control participants, there were about half as many participants with autism who gave spontaneous expressions of greeting, and a substantial proportion of those with autism failed to respond when the stranger said 'Hello'. Whereas all the 24 young people without autism made eye contact, and no fewer than 17 smiled, a third of those with autism never made eye contact and only six smiled. The results for the farewell episode were broadly similar. Half the individuals without autism but only three of those with autism made eye contact and said goodbye. Of these, nine of the non-autistic individuals but not a single autistic individual also smiled. Here it was especially interesting that participants with autism rarely gave a wave of farewell when the investigator called 'Goodbye' as the child left through the door, and each of the five who did so made strangely limp and often ill-directed flaps of the hand that were not coordinated with eye contact.

We also asked the raters of our videotapes to make the following judgement about the greeting episode: 'Over this period and prior to sitting down, to what degree did you feel that the child engaged with the stranger?' The categories of response were either strongly engaged, somewhat engaged, or hardly, if at all engaged. It turned out that different judges who made these ratings independently were in good

agreement with each other. The results were that 14 of the 24 non-autistic children were judged to be strongly engaged, and only two hardly, if at all, engaged. In contrast, only 2 of the 24 children with autism were judged to be strongly engaged and 13 of them seemed hardly, if at all, engaged. These judgements more or less (but not quite) mapped on to the behavioural ratings.

I am going to complement these descriptions of social interactions with one of a series of experimental studies we conducted on 'emotion recognition' among children with autism (others of which are described in Hobson 1993a, 2002). Weeks and Hobson (1987) tested 15 autistic and 15 non-autistic retarded subjects who were similar in chronological age, sex and verbal ability. Participants were given a task of sorting photographs of people to 'go with' one or other of a pair of target photographs showing the head and shoulders of individuals who contrasted in three, two or one of the following respects: sex, age, facial expression of emotion and the type of hat they were wearing. The features that were varied were kept simple: either a happy or a non-happy (rather glum) face, either a male or a female, either an adult or a child, either a floppy hat or a woollen hat. The instructions were that the child should notice one way in which the two pictures were different and sort 16 photographs of new people (with systematically varying characteristics) accordingly.

We adopted the following method. Suppose that a child initially sorted by hat. The next pair of target photographs would be of individuals wearing the same hat, but still different in emotional expression and sex. The same 16 sorting cards were shuffled, and subjects sorted them afresh. If the child then sorted by sex, the targets would be changed so that they now wore the same hat and were of the same sex. Finally, only one designated feature remained as a basis for sorting.

The results were that two-thirds of the subjects of each diagnostic group began by sorting the photographs according to sex, or more accurately, by some feature that corresponded with sex (which with the materials we were using might have been hair length). There was therefore no question that the children with autism could discriminate the photographs and sort accordingly. However, whilst 10 of the 15 non-autistic children sorted by emotional expression before they sorted by hats, only 3 of the 15 autistic children did so, and 9 children sorted by hats in preference to facial expression (the remaining three autistic subjects sorted by neither hat nor facial expression). Moreover, equal numbers of autistic and control subjects spontaneously sorted by sex at some point during the task and equal number sorted by hat; but whereas all 15 of the non-autistic children sorted by facial expression sooner or later, only 6 out of 15 children with autism did so. For the remaining nine children with autism, the contrast in facial expression did not seem to register. Even when given explicit instructions to sort by facial expression, 5 out of 15 participants with autism still failed to sort consistently.

Perhaps I should say why I cite these observations. Through Kanner's description of Charles, one can easily understand why a number of clinicians have described children with autism as treating other people as pieces of furniture – for the reason that the other people (including the clinicians) felt *themselves* to be treated as pieces of furniture. I stress this because it is a good example of the nature and implications

of intersubjectivity. We do not usually feel like pieces of furniture in the presence of people, for the reason that the relations between ourselves and others implicate mutual affective adjustment and exchange. We feel ourselves to be persons in the presence of a person – and note that this entails a form of understanding what a person is. Then through the Hello-Goodbye study, one can see how, although intersubjective engagement works through the coordination of bodily expressed feelings such as facial and other gestures, and such coordination is notable by its relative absence among children with autism, it is what we apprehend *through* such gestures that makes intersubjectivity or its relative absence so powerful. In addition, there was that curious and striking observation of the children's waves that were not really waves, either in communicative intent (as far as one could see – if anything, they felt more like unfocussed brushing away) or in their limp and unformed shape. Finally, the experiment with photographs of faces (for the results of which, incidentally, there is corroborating evidence in Jennings 1973) reveals that even under conditions that are far from life-like interpersonal exchanges, children with autism neither notice nor discriminate among facial expressions of emotion in the way that is characteristic of other children. What grabs *us* about other people, their emotional attitudes, is partly lost on children with autism. And note that being grabbed is a motivational as well as an emotional/cognitive matter.

What these studies of autism bring into relief, then, is a quality or level of social functioning that is characteristic of human beings who do not have autism. Of course, it requires additional evidence to justify claims that qualities of intersubjective engagement are developmentally prior to and foundational for more intellectually elaborated understandings of persons with minds, and that impairments in such forms of engagement are responsible for the limited understanding that children with autism acquire about the mental states of other people. On the matter of temporal precedence, I simply note there is abundant evidence not only that young typically developing infants show finely coordinated interpersonal relatedness, at least from the second month of life (and in some respects, before this), but also that they react strongly to events such as their mothers adopting a 'still-face' stance that disrupt the kinds of social interactions that they anticipate (Tronick et al. 1978). Correspondingly, there is now substantial evidence that as infants, many children with autism had abnormalities in mutually coordinated, affectively patterned relations with others (e.g. Charman et al. 1997; Wimpory et al. 2000). So it is reasonable to conclude that, whether or not such aspects of social relatedness are essential for acquiring a conceptual grasp that others have their own subjective attitudes and perspectives on the world, they are certainly candidates for this role.

3.1.3. Understanding persons

Here is what an intelligent adolescent with autism said to his physician, Cohen (1980, p. 388):

> I really didn't know there were people until I was seven years old. I then suddenly realised that there were people. But not like you do. I still have to remind myself that there are people… I never could have a friend. I really don't know what to do with other people, really.

Note that this person did not say he was unable to mind-read, although this, too, is not unknown as a complaint from someone with autism. What comes through in his description is his lack of the kinds of experience of people that make it unnecessary to remind oneself of their existence. What kinds of experience are those?

As I have mentioned already, Hamlyn (1974) argues that in order to understand what persons are, one needs to understand the kinds of relation that properly exist between persons. With good reason, Hamlyn stresses that those relations entail mutual feelings (also Hobson 1993b). I have been trying to develop a particular aspect of this account by illustrating how, through a person's perception of and responsiveness to the bodily expressed subjective states of others, that person's own subjective states are altered in mutually coordinated ways. We have biologically determined propensities to affect and be affected by others, and for this reason we become engaged with others psychologically. We do not need to 'infer' that people (or should that be bodies?) have minds.

Now the challenge is to characterize what such engagement entails, especially in the early phases of infancy, so we can envisage how the structure of such engagement provides a suitable starting point for developing concepts of mind.

I find it helpful to frame the problem in terms of connection and differentiation (Hobson 1993a; Werner and Kaplan 1963/1984). We need to be psychologically connected with other people in one-to-one interactions and in our ways of experiencing the world, because in due course we need to understand that we have commonality with others in being persons with minds, each one of whom is a self. At the same time, we need to grasp how we are differentiated from others, in the sense that different selves have different feelings, wishes, preferences, beliefs, personality characteristics, and so on. Early on in our lives, we come to know that everyone else has his or her own subjective experiences, as do we. Moreover, we live in a world on which each person has both a shared perspective, in that it is 'our' world, and different perspectives, in that each of us subsumes that world under a description-for-oneself. By the way, in the study of early development and developmental psychopathology, we need to distinguish not only among different forms of perspective, but also different forms of understanding perspectives. For example, it is one thing to understand one's own and others' intentions or preferences (where children with autism are relatively adept), and quite another to understand one's own and others' feelings, beliefs and knowledge; it is one thing to have sensory-motor-affective 'understanding'-in-action-and-feeling, and quite another to have conceptual understanding of what it is to entertain a perspective. Indeed, as I have argued elsewhere (Hobson 2005a), one should really set any reference to non-conceptual forms of 'understanding' in quotation marks, if one is to avoid theoretical slippage in characterizing the steps of development.

At the level of one-to-one interpersonal engagement, there is something especially important about the ability to share experiences, both from a phenomenological and developmental point of view (Hobson 1989; Hutto 2002). This is because it is the paradigmatic case of intersubjective connectedness and differentiation. True, it may be the case that very early in life, infants register that 'this' is special, where

'this' involves what *we* see as interpersonal coordination but the infant is unaware of anything other-person-centred in the experience. Yet from birth onwards, and especially in settings where the infant imitates another person (e.g. Field et al. 1982; Meltzoff and Moore 1977), there is also indication that the infant registers when another person's actions and probably attitudes differ from his or her own. Recall how 2-month-olds can be seen to react negatively to disruption in face-to-face communication with their mothers, and they often appear to make bids to the mother to reinstate reciprocal interaction. Most certainly by the end of the first year of life, when an infant makes efforts to share objects and events with others as one manifestation of newfound capacities for 'joint attention', the infant's awareness of such connectedness and differentiation is clearly manifest.

Now I believe the kinds of intersubjectivity achieved by humans are unique to humans, and that they account for critical differences between, say, chimpanzees and people in cognitive (not just social) development. So consider how empathy involves more than responding to someone emotionally: It also means responding to the other person's feelings *as* the other's feelings. In other words, the feelings involved in an empathic response are both one's own and experienced in relation to the subjective experiences 'felt' as the other's (Scheler 1954). The same applies to other forms of sharing of the world. True sharing involves movement towards and adoption of aspects of the other person's psychological stance vis-à-vis objects or events and coordination with one's own now-expanded subjective state (Hobson 1989). The important thing here is that one participates in the other person's state and maintains awareness of otherness in the person with whom one is sharing, whilst also being affectively involved from one's own standpoint.

Now to the crux: I suggest that we need to posit a primitive, biologically based mechanism of *identification* to explain how human forms of sharing are as they are and to account for the development of conceptual understanding of the nature of selves-with-minds (and with this, concepts of mental states). As a corollary, I suggest that in most but not all cases of autism, affected individuals have a weak propensity to identify with others. The 'not all' caveat is needed not only because of the heterogeneity of autism, but more specifically because certain conditions such as congenital blindness can render a child functionally unable to achieve what identification usually enables children to achieve, even though the children's potential to identify is more or less intact (see Hobson et al. 1997; Hobson 2005b, for discussion of how the high prevalence of autism among congenitally blind children bears upon the arguments of this chapter). I would stress that it is in being *moved to* the emotional stance of others, and therefore in adopting alternative person-centred perspectives, that children with autism are especially handicapped (Hobson 2002; Hutto 2003).

It is not easy to define identification, but one helpful definition is that provided by Laplanche and Pontalis (1973): 'Psychological process whereby the subject assimilates an aspect, property or attribute of the other and is transformed, wholly or partially, after the model the other provides.' The value of such a definition is that it highlights how one person assumes the behaviour or other characteristics of another

person. For other purposes, for example when applied to levels of interpersonal exchange that might be subject to experimental investigation (and one such study is coming up), it is appropriate to narrow the definition: To identify with someone else is to relate to the actions and attitudes of someone else from the other's perspective or stance, in such a way that a person assimilates the other's orientation towards the world, including towards the self, so that this orientation becomes a feature of the person's own psychological repertoire (Hobson 1993a, 2002; Hobson and Lee 1999). The important thing about identification is that one feels in accordance with the other, but one does not entirely become the other; and one can then make the other's feelings or attitudes one's own, whilst still entertaining those feelings or attitudes as separate from other aspects of the self. In this way the possibility arises that, in the course of development, one can come to punish oneself, be patient with oneself, and so on.

If this account is correct, then it carries implications for many qualities of human relations. For example, if it is the case that in the course of moment-by-moment social interactions, typically developing children are not only perceiving but also identifying with the bodily expressed movements, actions and gestures of others, and so being 'shaped' by mutual engagement, the effects of a relative lack of such processes in children with autism will have negative implications for the development of coordination of movement, action and gesture. Or at another level, if young children identify with the assertive agency of others or others' bids for possessing things or their aggressiveness or a range of other ways of relating, and at the same time retain awareness of their own relation to such patterns of relatedness, then this becomes a means to creating an 'internal world' of mutually interacting stances, each of which the children might assume. This is what psychoanalysts have been writing about for years.

3.1.4. Sharing en route to thinking: joint attention and imitation

When we move from the earliest months to the end of the first year of life, and consider how infants achieve joint attention and what has been called 'secondary intersubjectivity' with someone else (Trevarthen and Hubley 1978), we need to consider how not only sharing of subjective states, but also coordination of subjective states in relation to a shared external world, enter the picture.

At around the end of the first year of life, typically developing infants begin to follow others' points, often looking between the object indicated and the person's face, they offer and show things to others apparently to share them and for no instrumental purpose, they request things and respond to requests and begin to inform others (e.g. Carpenter et al. 1998). In addition, they begin to engage in social referencing: If they are faced with something like an anxiety-provoking but interesting object or event such as an active robotic toy, they will look to their caregiver's face and alter their attitude to the toy in response to the caregiver's attitude towards that same toy, but not if the adult is looking elsewhere (e.g. Hornik et al. 1987; Sorce et al. 1985; Walden and Ogan 1988). This is especially important, insofar as we observe the infant seeking, finding and being affectively

moved by another person's attitude to a common focus in the world. Although it would be premature to say that the infant conceptualizes another person as having a perspective, it would not be premature to say that the infant is assuming new psychological orientations *through* other people. And at the level of 'implicit understanding', the infant has grasped some rather important things about the way other people relate to the world.

In the case of autism, there are well-attested impairments in joint attention. Research by Mundy and Sigman and their colleagues (Mundy and Sigman 1989; Sigman et al. 1986) has indicated that joint attention, and especially the initiation of joint attention, reflects something very basic about the impairments in intersubjective engagement that characterize autism. Sigman et al. (1992) also demonstrated how young children with autism show markedly reduced social referencing. More recent investigations of ever younger groups of children with autism, either through direct study (e.g. Charman et al. 1997) or through parental report (e.g. Wimpory et al. 2000), have established that abnormalities in joint attention as well as affective responsiveness are early appearing in many of the children.

Another ability that flowers at the end of the first year of life in typically developing infants is that of imitating other people. For example, in a longitudinal study Carpenter et al. (1998) reported that it was around the time of their first birthdays that infants imitated an adult in both instrumental actions (e.g. opening a hinge) and arbitrary actions (e.g. patting the surface of a box). Imitation is especially important for my account, in that it *can be* a more or less direct manifestation of the process of one person identifying with another. I stress 'can be', because there are other ways that one person can copy another's actions. I am going to illustrate what I mean by returning to the case of autism.

It is a relatively consistent finding from the studies of imitation in children with autism that they display abnormalities in the ability to imitate facial, vocal and gestural expressions, as well as meaningless movements (e.g. DeMyer 1975; Smith and Bryson 1994). On the other hand, they are relatively adept in copying goal-directed actions (e.g. Charman and Baron-Cohen 1994), and importantly, they may become more socially engaged and interactive when an adult imitates their actions (Dawson and Adams 1984; Nadel et al. 1999). Then, as several authors have noted (e.g. Barresi and Moore 1996; Ohta 1987; Whiten and Brown 1998), children with autism sometimes appear to reproduce actions as seen when they watch a demonstrator, instead of reproducing what the demonstrator would have seen when acting. For example, Smith and Bryson (1998) described an error involving rotation of the hand by 180°, something that seemed to reflect a difficulty in translating the view of an action made by someone else into a matching action of the self. Striking evidence of such problems with self/other transposition, along with a marked lack of the propensity to imitate the style (mainly, harsh or gentle) with which a person demonstrated actions, was reported by Hobson and Lee (1999). For example, when one of the investigators put a pipe rack against his shoulder and strummed it harshly with a stick, most participants without autism imitated the *person* of the demonstrator by putting it against *their* shoulder and strumming

harshly. By contrast, most of those with autism copied the *action* of strumming but without a recognizable style, and with the pipe rack positioned on the table in front of them and not against their shoulder. And here it is worth recalling those limp, ill-directed waves we saw in the Hello-Goodbye study, because they appear to represent a naturally occurring instance of actions unshaped by identification with what others express and intend to communicate to someone else through movement and gesture.

We are now in a position to consider whether we might unearth more specific evidence to suggest that identification is operative both in episodes of joint attention that involve sharing of experiences, and instances of imitation. To recap: when, through identification, an individual shares experiences of the world with someone else in joint attention, he or she *both* resonates to the attitudes of the other from the others' bodily anchored stance *and* maintains enough of his or her own starting state to make the sharing 'sharing' (and not mere adjustment). When, through identification, an individual is motivated to imitate someone else, this entails that the individual adopts the other person's stance in behaving and relating to the world, rather than copying the action *per se*. For this reason, the individual tends to copy the style and self/other orientation of the other's actions, a process that promotes role-reversals in action and attitude.

We devised a fresh approach to investigating self/other orientation (Meyer and Hobson 2004; and see Hobson and Meyer 2005, for related findings from a semi-structured communicative 'sticker test'). We tested 32 children between the ages of six and 14 years, half with autism and half without autism but instead with learning difficulties or developmental delays. Our procedure began with the tester and child seated on carpet squares on the floor, situated on opposite ends of a testing mat, directly across from each other at a distance of approximately 20 in. There were four actions demonstrated, and each was presented in two different ways. For example, the tester picked up a small wheel with a metal handle (a castor from furniture) that had been placed in the middle of the testing mat, and proceeded to roll it from one side of the mat to the other, *either* across the mat directly in front of herself (i.e. from left to right, neither away from nor towards herself) *or*, leaning forward, across the front of the participant; another example was where she *either* lifted a blue box from its position in front of herself, placed it on top of a box positioned in front of the participant, and then returned it to the starting point, *or* lifted the box closest to the participant, placed that on top of the box nearest herself, then returned it to its original position. For each of the four conditions, children saw the investigator produce the action in one of two possible orientations – close to or towards herself, or close to or towards the child – on the first testing session, and saw the alternative orientation for each condition in a second session on another day. After demonstrating each action, the examiner returned the object(s) to their original positions and instructed the child: 'Now you.' There was no explicit instruction to copy what the investigator had done.

The children's subsequent actions were scored as reflecting *identification* if the child copied the investigator's stance (i.e. the action in relation to self or

other). For example, identification occurred when the children imitated the tester's close-to-self-orientation by rolling the wheel close to him/herself, or copied close-to-other-orientation by rolling the wheel close to the tester. In those cases where the action was *not* characterized by identification with the other person's stance, we made a further categorization: Was it simply that the response was out of keeping with identification, and perhaps without specific orientation, or did it take the form of an exact replication of the action, so that it resulted in a second run-through of the child's original view of what was done to the object(s)? In the latter case, we classified the response as being an instance of *geometric repetition* (akin to the 'reversal errors' seen in studies of gesture imitation).

As we predicted, the children with autism were significantly less likely to imitate the self-other orientation of the actions. While half of the children in the comparison group copied the self-other orientation of the actions on at least half of the eight trials, for example, only 3 out of the 16 children with autism did so; and from a complementary perspective, six of the participants with autism imitated self-other orientation on fewer than two occasions, while only one participant in the comparison group did so as infrequently as this.

When we examined the children's responses according to the most prevalent category of response, the strategy of geometric repetition – that is, responding so that the physical movements and locations of the objects acted-upon were replicated – was predominant among some (albeit a minority) of the children with autism, but none of the children in the control group. This result is reminiscent of instances of pronoun reversal, mimicry and echolalia reported in clinical accounts and early studies of autism, and of sporadic instances of 'reversal errors' in a number of recent studies on imitation (e.g. Smith and Bryson 1998). It might suggest that some of these children had a natural propensity to be object- or stimulus-bound in their focus of attention. In our view, it is more plausible that they had developed this mode of apprehending and/or dealing with the world because they relatively lacked an orientation towards other people's stance-in-acting.

Then in a critical further step, we tested a prediction that specifically sharing looks (reflecting identification) would have a positive relation to the propensity to imitate self/other orientation (Meyer and Hobson 2005). The demonstration and imitation sequences of the self/other orientation study were coded by an independent naïve judge (reliable with a second rater) for (a) direction of gaze – to the tester, object, or away and (b) quality of joint attention looks – sharing, checking, or orientating to the speech or movement of the tester. Results were that children with autism spent less than half as much time looking at the tester, and significantly more time looking at the objects, relative to children in the comparison group. This difference was not specific to a particular quality of joint attention look, as the pattern was similar for sharing, checking and orienting looks. Furthermore, the percentage of time spent looking at the tester overall, as well as frequency of checking and orienting joint attention looks, was *not* related to imitation of self-other orientation in either group. By contrast, and as we had predicted, *sharing*

looks were specifically and significantly associated with the children's propensity to imitate self-other orientation, both within and across the two groups.

Here is evidence that a particular kind of involvement with another person, one reflected in sharing looks, may be associated with a particular kind of imitation, namely that involving the kind of self/other transposition that often, we believe, reflects the process of identification. Or to put this in simpler terms, if you are engaging with someone else, you will probably be assuming their stance (as part of a more complex relation); if you are not, you probably won't.

Yet to assume a stance is not the same as understanding what it is to assume a stance, and this is a critical achievement of the second year of life.

3.1.5. Mind, self and symbol

We are nearing the end of the chapter, and I have only reached the half-way stage in accounting for the emergence of conceptual thought. But so be it: the very earliest period of development is critical for setting the stage for what follows.

There are very important changes that occur in typically developing children's relations with others, just as there are in their language and thinking, around the middle of the second year of life. In my view, the emergence of new levels of self- and other-awareness and the appearance of symbolic play, the vocabulary spurt and other cognitive achievements that have sometimes been gathered under the term 'metaprepresentation' (Leslie 1987) all pivot around an insight: That one is a self like other selves, with each self having (amongst other things) a mental orientation of one's own that can be communicated to others.

In order to achieve this insight, children need to have had experiences that enable them to grasp two things: Firstly, the distinction between objects and events on the one hand, and people's attitudes or takes on those same objects and events on the other; and secondly, that they can apply person-anchored meanings to new objects so that these serve as symbolic vehicles. This is rather a concrete way to express the matter, because children's awareness of symbolic meanings both arises from and applies to a range of social-communicative events. Moreover, mental representations of symbolic vehicles comprise the fabric of thought (note: an outside-in development, as Kaye 1982, expressed it). Anyway, all this entails something else, namely that children conceptualize that each person, including the children themselves, is a self with his or her own orientation towards the world.

There is a paradox here. In order to understand what it is to entertain a perspective, which is what is at stake, and in order to conceive what it means to be a self who has a perspective, the child needs the conceptual equipment (symbols) to achieve these higher order, more or less explicit forms of understanding; but in order to employ symbols in such a way that such understanding is possible, for example in articulating subject-predicate relations, the child needs to be aware of him- or herself as a self who can grasp and apply perspectives. Perhaps this last claim is too strong, because although a child *does* need to know what he or she is doing in choosing to apply new meanings in symbolic play, the child may need but a partial grasp of such mental activity when it comes to using symbols in language (which

anyway involves a cluster of different pragmatic devices). Whatever the case here, the paradox is resolved if we appreciate that the synchrony in developing these seemingly separate cognitive abilities is no coincidence. Conceptualizing self and other, and grasping what it means to symbolize, are two sides of the same coin.

Elsewhere I have attempted to say more about how these new forms of understanding crystallize out of the preceding forms of 'understanding' that I have been trying to characterize (especially Hobson 1993a, 2000). Of particular importance is how the child identifies with others not only in relation to a shared, multiply referenced world, but also in relation to the child's own bodily anchored actions and attitudes. This is important for the reason that the child needs to have a route through which he or she *relates to his or her relations with the world*, and that route is by way of identifying with other people relating to what the child is doing and feeling. Much of this shifting among perspectives happens in mutually coordinated communication with others, of course.

For the present purposes, I have highlighted two things. The first is how, prior to and as a precondition of *thinking* about perspectives, the child is drawn into *assuming* multiple perspectives on shared objects and events, and through the structure of identification, bringing more than one perspective in relation to another within his or her mind. This creates conditions in which the child can recognize that particular objects and events can be the focus of alternative perspectives, and therefore that meanings-for-persons are separable from the objects and events to which those meanings usually apply. The second thing is that there is a person-anchored quality to these perspectives, so that the child's engagement with the attitudes of others creates conditions in which the child can recognize not just that there are perspectives, but that *people* have perspectives.

In outline, then, the story goes as follows. With dyadic intersubjective engagement already well-established, around 9 months of age infants begin to apprehend how a single object or event can be the target of different person-anchored attitudes. A next step is when, by the middle of the second year of life, they not only apprehend, but also conceptualize how people have different self-anchored perspectives. Then by four years, they achieve a new understanding of what 'reality' means, namely the supra-personal characterization of the way things are, a state of affairs that transcends individual subjective perspectives in the sense that anyone in the right position would assent to this particular way of looking at things. This amounts to a change in their understanding of what it means to know or believe to be true of reality. All this presupposes the ability to share experiences with others, to fix symbolic vehicles (including words) through communication with others, and to arrive at those agreements in judgement (Wittgenstein 1958) that underpin conceptual development. Sharing, communicating and agreeing with others means having the kinds of relations with others that entail intersubjective connectedness and differentiation. It seems to me that such relations require the kind of structure introduced into human social life by the process of identification.

Yet it remains to find a place in this picture for all those qualities of interpersonal relatedness I referred to earlier. Why are possessiveness, competing, and so on

important, beyond giving content to knowledge of people? Well, that is no small matter: through identifying with just such attitudes one gets under another person's skin, as it were, and into the experience of that person relating to the world with possessiveness, rivalry, aggressiveness, compassion, and so on. These modes of understanding, broadly speaking empathic in quality, are what gives body (not just content) to our knowledge of others. Beyond this, however, such relations are important because they are the stuff with which identification works. It is in feeling the other's possessiveness, for example, that my experience of self-other relations is reinforced. One effect of such identification with a range of attitudes is to orientate a child (or adult) to what I earlier referred to as the person-anchored nature of perspectives.

I shall allow Gerhard Bosch (1970), writing on autism, to bring home the point. Through clinical descriptions, Bosch conveys how children with autism often seem to lack a sense of possessiveness as well as self-consciousness and shame, to be delayed in 'acting' on others by demanding or ordering, and to be missing something of the "'self-involvement', the acting with, and the identification with the acting person" (p. 81). He also suggests that "counter-attack or defense is impossible because the child has no experience of an attacking or defensive relationship with others" (p. 99). Perhaps most prescient of all, Bosch emphasizes that "delay occurs in the constituting of the other person in whose place I can put myself… [and]… in the constituting of a common sphere of existence, in which things do not simply refer to me but also to others" (p. 89). He does not discount the fact that many children with autism do develop relationships that 'without doubt may be termed personal' (p. 93); but when it comes to the kind of reciprocal relationship in which others' expressiveness is responded to as such, or where the child needs to establish a world as shared and at the same time experienced differently by others, profound difficulties arise.

These writings capture how identification operates. It is primarily an emotional business, both in its developmental underpinnings and as it applies across the lifespan. It shapes the field of the interpersonal and creates the domain of self in relation to other.

3.1.6. Simulation, theory theory, and where propositional attitudes come from

Suppose the account I have presented is more or less valid. What does it say about the origins and nature of interpersonal understanding, or folk psychology? In part, identification operates as a particular form of simulation, but in such a way as to provide a grounding for what subsequently becomes the ability to (knowingly) put oneself in someone else's shoes, and for imagining possible takes on the world. In part, too, identification provides what is necessary for theorizing – or better, a structure of experiencing others that is required for us to *know* (not theorize) that other people have minds, and to conceptualize the kinds of mental orientations or perspectives that belong to selves, both oneself and others. In addition, through identification we are 'moved' to adopt or otherwise respond to other-centred

perspectives, and this moving among perspectives, understood as such, becomes a feature of the individual's creative imagination.

There is one final matter I should like to address. This concerns the nature of propositional attitudes of the kind that are special to humans ('I believe that p', 'I desire that p', and so on). If we want to understand what such propositional attitudes are, and how they operate, we would do well to study how they are developed or constructed. According to the account I have given, to have these kinds of propositional attitudes presupposes understanding of what it means to have them. The reason is that they entail a particular species of psychological relation between a self and the world as falling under a particular description *for* that self, and the requisite form of relation needs to be represented within a child's mind. Essential to this mental relation is a grasp of the way in which others have attitudes to the world that can be characterized according to descriptions-for-persons, i.e. people have perspectively configured mental contents that could be cast in the form of propositions, or anchored in symbols. By 18 months of age, children have come to understand something (not everything) of this relation between persons' perspectives and the objects and events that they encounter, on the basis of the interpersonal coordination of attitudes as these attitudes are directed towards a shared, visually specified world. It is through such understanding that children acquire the capacity to shift among attitudes to propositional contents (since these are now represented as such), and among those contents themselves. Then as the years tick by, children come to understand new kinds of attitude that selves can assume not only towards the world, but also towards 'reality' as newly conceived: People can believe such-and-such to be *true of reality*, but they may be wrong. This account, then, is one that begins with an infant having and responding to people's attitudes, and takes a social-developmental route to explain how by middle childhood, children come to have the ability to think about such contents of other minds as beliefs.

Insofar as this account is true, understanding others is acquired through our engagement with their subjective attitudes. At each stage in early development, there is a growing differentiation of meanings-for-me from meanings-for-other. Thereby, mental space opens out as the child comes to relate to his or her own relations with the world.

This developmental story presents us with a particular view of the structure of thinking, or perhaps more strictly, a specifically human form of thinking. Thought is distilled out of affective relatedness, so representations still retain their connections with what they represent; propositions are distilled out of attitudes as these are given symbolic (often linguistic) anchorage, so they, too, remain linked to whatever they are supposed to be propositions about (Hobson 2000). Such psychological differentiation is possible because of a typically developing child's 'natural reactions of person to person' (Hamlyn 1974), that is, his or her engagement with the bodily expressed attitudes of others towards the world and the child him- or herself.

The upshot of my argument, I suppose, is that much of contemporary theorizing about the nature of human cognitive functioning is misleadingly computational, individualistic and overly focused on the prediction and explanation of behaviour.

We come to think because we are engaged with other embodied people, and engaged in such a way that we can incorporate other-centred psychological perspectives in our minds, whilst achieving awareness of our own view as our own. We are not given mental architecture for entertaining propositional attitudes, any more than we are faced with the need to infer the existence of mental states in others; rather, we come into the world with attitudes that are relational vis-à-vis the social and non-social worlds, respectively, and we need experience of both people and things in order to achieve the differentiation of thinking from the objects of thought. Vygotsky (1962, p. 8) captured something of this when he wrote: '… every idea contains a transmuted affective attitude toward the bit of reality to which it refers'. I would add that every thought also manifests structure that is the product of social coordination in attitudes towards the world. We share, therefore we think.

ACKNOWLEDGEMENTS

This chapter is being written while I am a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford. I am very grateful both to the Center and to the Tavistock and Portman NHS Trust for giving me the chance to be here with stimulating colleagues. I thank my colleague Dr Jessica Meyer for her many contributions to the work.

REFERENCES

Andrews K (this volume) Critter psychology: on the possibility of nonhuman animal folk psychology.

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind"? Cognition 21:37–46.

Barresi J, Moore C (1996) Intentional relations and social understanding. Behav Brain Sci 19:107–154.

Bosch G. (1970) Infantile autism. Transl. by Jordan D, Jordan I. Springer, New York.

Carpenter M, Nagell K, Tomasello M (1998) Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monogr Soc Res Child Dev Serial no. 255 63:4.

Charman T, Baron-Cohen S (1994) Another look at imitation in autism. Dev Psychopathol 6:403–413.

Charman T, Swettenham J, Baron-Cohen S, Cox A, Baird G, Drew A (1997) Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. Dev Psychol 33:781–789.

Cohen DJ (1980) The pathology of the self in primary childhood autism and Gilles de la Tourette syndrome. Psychiatr Clin North Am 3:383–402.

Dawson G, Adams A (1984) Imitation and social responsiveness in autistic children. J Abnorm Child Psychol 12:209–226.

DeMyer MK (1975) The nature of the neuropsychological disability in autistic children. J Autism Child Schizophr 5:109–128.

Field TM, Woodson R, Greenberg R, Cohen D (1982) Discrimination and imitation of facial expressions by neonates. Science 218:179–181.

Gallagher S (2007) Logical and phenomenological arguments against simulation theory. In: Hutto D, Ratcliffe M (eds) Folk psychology re-assessed. pp 63–78.

Gopnik A, Meltzoff A (1997) Words, thoughts, and theories. MIT Press, Cambridge, MA.

Hamlyn DW (1974) Person-perception and our understanding of others. In: Mischel T (ed) Understanding other persons. Blackwell, Oxford, pp 1–36.

Hamlyn DW (1978) Experience and the growth of understanding. Routledge & Kegan Paul, London.

Hobson RP (1989) On sharing experiences. Dev Psychopathol 1:197–203.

Hobson RP (1990) On the origins of self and the case of autism. Dev Psychopathol 2:163–181.

Hobson RP (1991) Against the theory of 'Theory of Mind'. Br J Dev Psychol 9:33–51.
Hobson RP (1993a) Autism and the development of mind. Erlbaum, Hove, Sussex.
Hobson RP (1993b) The emotional origins of social understanding. Philos Psychol 6:227–249.
Hobson RP (2000) The grounding of symbols: a social-developmental account. In: Mitchell P, Riggs KJ (eds) Reasoning and the mind. Psychology Press, Hove, UK, pp 11–35.
Hobson RP (2002) The cradle of thought. Macmillan, London.
Hobson RP (2005a) Social engagement and understanding in chimpanzees and humans. Commentary on Tomasello M, Carpenter M. The emergence of social cognition in three young chimpanzees. Monogr Soc Res Child Dev, Serial no. 279 70:133–152.
Hobson RP (2005b) Why connect? On the relation between blindness and autism. In: Pring L (ed) Autism and blindness. Whurr, London, pp 10–25.
Hobson RP, Lee A (1998) Hello and goodbye: a study of social engagement in autism. J Autism Dev Disord 28:117–126.
Hobson RP, Lee A (1999) Imitation and identification in autism. J Child Psychol Psychiatry 40: 649–659.
Hobson RP, Meyer J (2005) Foundations for self and other: A study in autism. Dev Sci 8: 481–491.
Hobson RP, Brown R, Minter M, Lee A (1997) 'Autism' revisited: the case of congenital blindness. In: Lewis V, Collis GM (eds), Blindness and psychological development in young children. British Psychological Society, Leicester, UK, pp 99–115.
Hornik R, Risenhoover N, Gunnar M (1987) The effects of maternal positive, neutral and negative affective communications on infant responses to new toys. Child Dev 58:937–944.
Hutto DD (2002) The world is not enough: shared emotions and other minds. In: Goldie P (ed) Understanding emotions: mind and morals. Ashgate, Burlington, VT.
Hutto DD (2003) Folk psychological narratives and the case of autism. Philos Pap 32:345–361.
Hutto DD (2004) The limits of spectatorial folk psychology. Mind Lang 19:548–573.
Jennings WB (1973) A study of the preference for affective cues in autistic children. Unpublished doctoral dissertation, Memphis State University.
Kanner L (1943) Autistic disturbances of affective contact. Nerv Child 2:217–250.
Kaye K (1982) The mental and social life of babies. Methuen, London.
Laplanche J, Pontalis J-B (1973) The language of psycho-analysis. Hogarth, London.
Leslie AM (1987) Pretense and representation: the origins of "theory of mind". Psychol Rev 94:412–426.
Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. Science 198:75–78.
Meyer JA, Hobson RP (2004) Orientation in relation to self and other: the case of autism. Interact Stud 5:221–244.
Meyer JA, Hobson RP (2005) Identification: the missing link between joint attention and autism? Dev Psychol (in press).
Mundy P, Sigman M (1989) The theoretical implication of joint-attention deficits in autism. Dev Psychopathol 1:173–183.
Nadel J, Guerini C, Peze A, Rivet C (1999) The evolving nature of imitation as a format for communication. In: Nadel J, Butterworth G (eds) Imitation in infancy. Cambridge University Press, Cambridge, pp 209–234.
Ohta M (1987) Cognitive disorders of infantile autism: a study employing the WISC, spatial relationship conceptualization, and gesture imitations. J Autism Dev Disord 17:45–62.
Piaget J (1972) The principles of genetic epistemology. Transl. by Mays W. Routledge & Kegan Paul, London.
Scheler M (1954) The nature of sympathy. Transl. by Heath P. Routledge & Kegan Paul, London.
Sigman M, Mundy P, Sherman T, Ungerer JA (1986) Social interactions of autistic, mentally retarded and normal children and their caregivers. J Child Psychol Psychiatry 27:647–656.
Sigman MD, Kasari C, Kwon JH, Yirmiya N (1992) Responses to the negative emotions of others by autistic, mentally retarded, and normal children. Child Dev 63:796–807.
Smith IM, Bryson SE (1994) Imitation and action in autism: a critical review. Psychol Bull 116:259–273.

Smith IM, Bryson SE (1998) Gesture imitation in autism I: nonsymbolic postures and sequences. Cogn Neuropsychol 15:747–770.

Sorce JF, Emde RN, Campos J, Klinnert MD (1985) Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. Dev Psychol 21:195–200.

Stawarska B (2007). Persons, pronouns, and perspectives: linguistic and developmental contributions to dialogical phenomenology. In: Hutto D, Ratcliffe M (eds) Folk psychology re-assessed. pp 79–99.

Strawson PF (1962, originally 1958) Persons. In: Chappell VC (ed) The philosophy of mind. Prentice-Hall, Englewood Cliffs, NJ, pp 127–146.

Trevarthen C, Hubley P (1978) Secondary intersubjectivity: confidence, confiding and acts of meaning in the first year. In: Lock A (ed) Action, gesture and symbol: the emergence of language. Academic Press, London, pp 183–229.

Tronick, E, Als H, Adamson L, Wise S, Brazelton TB (1978) The infant's response to entrapment between contradictory messages in face-to-face interaction. J Am Acad Child Adolesc Psychiatry 17:1–13.

Vygotsky LS (1962) Thought and language. Transl. by Hanfmann E, Vakar G. MIT Press, Cambridge, MA

Walden TA, Ogan TA (1988) The development of social referencing. Child Dev 59:1230–1240.

Weeks SJ, Hobson RP (1987) The salience of facial expression for autistic children. J Child Psychol Psychiatry 28:137–152.

Werner H, Kaplan B (1963/1984) Symbol formation. Erlbaum, Hillsdale, NJ.

Whiten A, Brown J (1998) Imitation and the reading of other minds: perspectives from the study of autism, normal children, and non-human primates. In: Braten S (ed) Intersubjective communication and emotion in ontogeny. Cambridge University Press, Cambridge, pp 260–280.

Wimpory DC, Hobson RP, Williams JMG, Nash S (2000) Are infants with autism socially engaged? A study of recent retrospective parental reports. J Autism Dev Disord 30:525–536.

Wittgenstein L (1958) Philosophical investigations. Transl. by Anscombe GEM. Blackwell, Oxford.

Zahavi D (2007) Expression and empathy. In: Hutto D, Ratcliffe M (eds) Folk psychology re-assessed pp 25–40.

4. LOGICAL AND PHENOMENOLOGICAL ARGUMENTS AGAINST SIMULATION THEORY

Theory theorists conceive of social cognition as a theoretical and observational enterprise rather than a practical and interactive one. According to them, we do our best to explain other people's actions and mental experience by appealing to folk psychology as a kind of rule book that serves to guide our observations through our puzzling encounters with others. Seemingly, for them, most of our encounters count as puzzling, and other people are always in need of explanation. By contrast, simulation theorists do their best to avoid the theoretical stance by using their own experience as the measure of everyone else's. When it comes to explaining how we understand other people some of the very best contemporary philosophers, psychologists, and neuroscientists are simulationists. For example, Vittorio Gallese, Alvin Goldman, Robert Gordon, Jane Heal, Susan Hurley, and Marc Jeannerod. This short list of simulationists, however, already involves some problems. Not everyone on this list understands simulation in the same way. In effect, there are different simulation theories, and although it is important to distinguish them, and I will do so before I go much further, I will in the end argue against all of them. For several reasons, I don't think that the concept of simulation explains our primary and pervasive way of understanding others, any more than theory theory (TT) does.

It should be obvious from what I've just said that being against simulation theory (ST) in this way does not mean that I favor its main competitor, TT. The alternative to both of these approaches is not something that I will directly try to argue for here. I've done so elsewhere (Gallagher 2001, 2004, 2006). But part of the argument that I develop against simulation here feeds into and depends on that alternative. So a quick summary is in order. Here it is. There is good evidence from developmental psychology, neuroscience, and phenomenology that we understand others, primarily and pervasively, in a nonmentalizing way, which is enhanced around the age of four years by a developing narrative competency. This doesn't rule out the possibility that in rare cases we do take a theoretical stance, or that we sometimes use simulation routines to solve puzzling cases. But these are the rare cases. Ordinarily, in everyday encounters, and in the pragmatic and social contexts that characterize our everyday encounters with others, we perceive their movements, gestures, facial expressions, and speech acts as meaningful and intentional, without worrying about their minds or about how to explain or predict their actions. If, instead of directly interacting with others, we are called upon (or we call upon ourselves) to think more deeply about them, our tendency is not to theorize about them, or to automatically put ourselves in their place, in the manner of a simulation,

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 63–78.

but to generate a narrative framework that would facilitate our understanding of them (Hutto 2003, 2004; Gallagher 2006; Gallagher and Hutto, in press).

Let me start the argument against ST by outlining several different versions of ST. Simulation theory, as an approach to theory of mind, has been developed in at least three different versions. An argument against one version will not necessarily work against all of them. So it is important to identify the specific targets in order to line the arguments up in the right way. The three different versions of ST are based on three different conceptions of what constitutes the simulation involved. In the first version, the simulation is something explicit, that is, an exercise of conscious imagination and deliberative inference. Goldman is a good representative of this position. In the second version, the simulation is also explicit, but does not involve a deliberative inference. This is Gordon's notion of radical simulation. In the third version, the simulation is something subpersonal, either a functional mechanism that is cashed out in terms of neuronal processes (e.g., Hurley 2005), or these neuronal processes themselves (e.g., Gallese 2001). There are, of course, hybrid theories that combine, within ST, explicit and implicit processes (e.g., Gallese and Goldman 1998), and hybrid theories that combine some version of ST with TT (see Currie and Ravenscroft 2002; Nichols and Stich 2003; Mitchell 2005).

4.1. EXPLICIT SIMULATION AND SOME ARGUMENTS AGAINST IT

In the explicit version of ST, simulation involves conscious or introspective mental states in which I imagine myself in the other's situation and use the model (simulation) that is generated to predict the other's mental states. Goldman, for example, argues that simulation is explicit insofar as it involves a conscious introspective use of the imagination to conceptually manipulate propositional attitudes (beliefs, desires). "When a mindreader tries to predict or retrodict someone else's mental state by simulation, she uses pretense or imagination to put herself in the target's 'shoes' and generate the target state" (Goldman 2005a). According to Goldman, simulation involves three steps.

> First, the attributor creates in herself pretend states intended to match those of the target. In other words, the attributor attempts to put herself in the target's 'mental shoes.' The second step is to feed these initial pretend states [e.g., beliefs] into some mechanism of the attributor's own psychology … and allow that mechanism to operate on the pretend states so as to generate one or more new states [e.g., decisions]. Third, the attributor assigns the output state to the target … [e.g., we infer or project the decision to the other's mind] (Goldman 2005b, pp. 80–81).

One might think that there would be a problem with the very first step: "the attributor creates in herself pretend states intended to match those of the target." This suggests that the simulator already has some idea of what's going on with the other person. The question then is where does such knowledge come from and why isn't that already the very thing we are trying to explain. Hybrid theorists who combine TT and ST suggest that folk psychology provides, not a sense of what is going on with the other person, but some general rules about how people think and behave in certain situations, and that this is what the simulationist can use to

generate the pretend mental states needed for the simulation process (e.g., Currie and Ravenscroft 2002). In contrast, Goldman appeals to subpersonal mirror resonance processes (discussed below), although he then faces the problem of how to translate these processes into a conceptual grasp of propositional attitudes. He proposes to solve the latter problem by an appeal to phenomenal properties of propositional attitudes (2002). One possibility is that a belief feels different from a desire because it is generated by different subpersonal processes, which are themselves generated by differential activations induced by our perception of the other. But, for purposes of simulation, it is clearly not the phenomenal properties that are important; the particular content of the beliefs or desires is what counts.

There are both logical and phenomenological arguments that have been made against this explicit version of ST. First, we should note that explicit ST is similar to an older theory called the argument by analogy. Gordon and Cruz (2003) recognize this forerunner as such. In the first half of the twentieth century arguments were raised against this theory by a number of philosophers, including Max Scheler and Gilbert Ryle, and some of the same arguments work against ST. Ryle, for example, argued that the logic of simulation isn't correct because the idea of imputing to a variety of others what is true of my simulated action ignores the diversity of their actions. "[T]he observed appearances and actions of people differ very markedly, so the imputation to them of inner processes closely matching [one's own or] one another would be actually contrary to the evidence" (Ryle 1949, p. 54). A similar objection to the logic of simulation was raised by Scheler (1954). If I project the results of my own simulation on to the other, I understand only myself in that other's situation, but I don't understand the other.[1] Given the diversity of motives, beliefs, desires, and behaviors in the world, it is not clear why such a simulation process would be at all reliable. Scheler also suggests another argument against ST. The explicit simulation process seems cognitively too complicated. Infants, and perhaps even animals, seem capable of understanding the intentions of others, but it would be difficult to attribute the complex cognitive processes involved in simulation to them.[2]

Here I'll propose one more argument, which I will call the simple phenomenological argument, against explicit ST. On the explicit version of ST, simulation is not only explicit but pervasive. That is, we use it all the time, or at least it is the default way of understanding others. Goldman (2002, pp. 7–8) thinks this is a moderate claim.

> The strongest form of ST would say that all cases of (third-person) mentalization employ simulation. A moderate version would say, for example, that simulation is the *default* method of mentalization … I am attracted to the moderate version …. Simulation is the primitive, root form of interpersonal mentalization.[3]

If simulation is both explicit and pervasive, then one should have some awareness of the different steps that one goes through as one consciously simulates the other's mental states. But there is no phenomenological evidence for this. When I interact with or come to understand another person, there is no experiential evidence that I use such conscious (imaginative, introspective) simulation routines. That is, when

we consult our own common experience of how we understand others, we don't find such processes. Of course, this is not to say that we never use simulations, but that in itself is telling. It may be the case that confronted with some strange or unaccountable behavior I do try to understand the other person by running a simulation routine (or by appealing to theory). I think this is the rare case, however. Moreover, it tends to stand out in its rarity. I can easily become aware that I am in fact taking this approach, and it is all the more apparent when I do this, simply because it tends to be the exception. But this tells against the idea that I employ simulation in the usual everyday circumstance. Most of our encounters are not third-person puzzles solved by first-person procedures. They are second-person interactions in which I easily have a sense of what is going on with the other person based on our common pragmatic or socially contextualized interactions, with no cognitive simulation required.

A possible defense of explicit ST is to make it a little implicit. Perhaps explicit simulation can be made so habitual that it becomes implicit, so that we do it without being aware that we do it, in the same way that we drive a car without being explicitly aware of all our driving habits, or in the same way that an expert may employ cognitive strategies that become so habitual that the expert is no longer aware of how she does what she does. The simple phenomenological objection would be that if such implicit processes stay at the personal level, they would remain accessible to conscious reflection, or at least they would become apparent, as unworkable habits, in problematic situations when our habitual strategies break down. We can become aware of a habit that we are not usually aware of in such circumstances. This simply does not seem to be the case for the sort of simulation process described by explicit ST. Indeed, we may find ourselves initiating simulation processes (or in some cases appealing to folk psychology) precisely in cases where our ordinary abilities to understand others break down.

4.2. RADICAL SIMULATION

Robert Gordon's notion of simulation is radical in the sense that, on his view, the capacity for simulation is what allows us, in the first place, to recognize another person as someone who is "mind-endowed" (Gordon 2004, p. 2). It is not simply that we understand the other to have beliefs and desires, and then use simulation to ascertain those mental states. Rather, simulation is the process that leads us to understand that the other is minded. At the same time, radical simulation is less mentalistic than the traditional explicit ST. I do not retreat introspectively to my own mind to run simulation routines by manipulating propositional attitudes like beliefs and desires. Rather, I put myself in the other person's perspective and look to see what she thinks is true about the world. This involves a transformation that takes place on the personal level. By using my imagination, I imagine/simulate what the other person must think in her situation. I do not imagine myself in her situation; I imagine *her* in her situation, by imaginatively occupying her situation.

The transformation involves an egocentric shift, but does not involve either introspection of my own mental states, or inference making about the other's mental states. I'm not concerned with mental states at all. I imagine, in the first-person, how the other person sees the world.

> The point I am making is that once a personal transformation has been accomplished, there is no remaining task of mentally transferring a state from one person to another, no question of comparing [the other person] to myself. For insofar as I have recentered my egocentric map on [the other person], I am not considering what [I] would do, think, want, and feel in the situation (Gordon 1995a, p. 54).

How do we effect this transformation? How do we know what the other person sees or thinks?

> My own view […] is that the method we ordinarily use is limited to identifying states in the first person, but, thanks to our capacity for imaginatively transforming ourselves into other "first persons," it is not exclusively a one-person method (p. 58).

We use our imagination and we use ascent routines. When, for example, I am asked whether I believe that the car is red, I don't start an introspective routine in order to find a belief located some place in my mind; rather, I look at the car to see if it is red, or I remember the car's color. I answer the question about my belief by appealing to a more primary cognitive procedure – perception or memory (see Evans 1982). In the same way, when I am asked whether my neighbor believes the car is red, I look at the car from her first-person perspective (Gordon 1995a, b). I settle the question of whether the other person believes P by asking, within the framework of the simulation, whether it is the case that P. I concern myself with the world from the other person's perspective.

The description of such ascent routines, which radically discounts mentalism and the propositional attitudes, is clearly a more parsimonious phenomenological account of our epistemological activities. It is, however, more difficult to cash out simulative transformations; and since these transformations are meant to be as explicit as the simulation routines described by Goldman, one should be able to get a purchase on them using phenomenological currency. Thus, the simple phenomenological argument works against radical ST also. Although Gordon does away with the need for an extra step involving inference, because we are "already there" in the other's perspective, these transformations still require an "as if" component. Otherwise, my own first-person perspective on the world would simply collapse into the first-person perspective of the other and the self/nonself distinction would disappear.

In ordinary circumstances, however, my experience of the other, even in the observational mode that seems to dominate theory-of-mind discussions, is not characterized by either an "as if" component or by a collapse of the self/nonself distinction. When I look out my window and see a man standing across the road, I need nothing more than my own perception of the environment to understand that his perspective is an opening onto just the same environment that I see, and I need not go further than that in most cases. That is, I don't have to transform myself into his perspective to know that he happens to see the road from an angle that differs

from my view. I can see that this must be the case simply from the differences that define our positions vis-a-vis the road and from the orientation and postural stance of his body. Ordinarily, it is enough to say that he sees the same road. And as for asking what might be going on inside the man's head, Gordon is right that I usually have no reason to ask. Even if the man suddenly starts jumping up and down in an excited fashion as he is looking and pointing up the road, I do not have to think about what propositional attitudes he might be entertaining, since I can see that he is excited about something. However, I do not attempt (or need to attempt) to put myself in his place in order to find out what might be exciting him; I do not attempt an egocentric shift of my perspective. In fact, I'm not quite sure what I could learn by doing so. Rather, I move to a position from which I can look up the road to see what he is seeing. No simulation is required, and if there is a transformation involved, it is simply the transformation of my own physical movement. I come to an understanding of his action, and his perception, and his excitement, through an ascent routine that is focused on our mutually perceived world, rather than on what might be happening within his mind, or within his perspective. Of course this may fail. If I see nothing up the road that might be the occasion of his excitement, then indeed, I may have to use some simulative or theoretical routines to work out what the man is doing. This won't get me very far, however, unless I gain some further information. And without further information, simply by transforming my egocentric perspective into his I will remain puzzled. Perhaps, by simulation, I would hypothesize that he is playing a joke on me, or, by appeal to theory, that he is delusional. But I would still need more information about the man's character – I would need to know the man's story – to determine whether my simulative or theoretical supposition was correct. This kind of strategy, however, is motivated only in unusual circumstances when the other person's behavior is puzzling. Most of my interactions with others are not like this and are not observational in nature. Most are second-person interactions where pragmatic and social contexts are shared; and in cases where I know the person, I may already have some sense of his character and possess a workable narrative framework to allow understanding without the mental acrobatics required for simulation routines or transformations.

4.3. IMPLICIT SIMULATION

ST can easily counter the simple phenomenological argument by moving to the more serious version of implicit simulation.[4] ST has gained more ground in recent years by appealing to good neuroscientific evidence involving subpersonal activation of mirror neurons, shared representations, or more generally, resonance systems. If simulation is subpersonal, and not something of which we would be aware, then phenomenology is not in a position to raise objections, since phenomenology gives us access only to conscious experience.

First, let's review the recent neuroscience. The basic finding in this regard is that one's motor system reverberates or resonates in one's encounters with others. My motor system is activated when I perceive another person performing an

intentional action, for example. Mirror neurons in the premotor cortex and in Broca's area of the human brain are activated both when the subject engages in specific instrumental actions, and when the subject observes someone else engage in those actions (Rizzolatti et al. 1996, 2000; Fadiga et al. 1995). Also, specific overlapping neural areas (shared representations), in parts of the frontal and parietal cortexes, are activated under the following conditions: (1) when I engage in intentional actions, (2) when I observe some other person engage in that action, and (3) when I imagine myself or another person engage in that action (e.g., Grèzes and Decety 2001). These subpersonal mechanisms are said to constitute a simulation of the other's intentions (Gallese 2001; Gallese and Goldman 1998). Gallese (2001, pp. 38–39) captures it clearly in his claim that

> when we observe goal-related behaviours … specific sectors of our premotor cortex become active. These cortical sectors are those same sectors that are active when we actually perform the same actions. In other words, when we observe actions performed by other individuals our motor system 'resonates' along with that of the observed agent …. action understanding heavily relies on a neural mechanism that matches, in the same neuronal substrate, the observed behaviour with the one [the observer could execute] …

The processes themselves are very real, and there is good neuroscientific evidence to support this. But is it appropriate to characterize these processes as simulations, as Gallese (p. 39) goes on to do?

> According to this hypothesis, 'understanding' is achieved by modeling [simulating] a *behaviour* as an *action* with the help of a motor equivalence between what the others do and what the observer does.

This is a subpersonal process generated by "automatic, implicit, and nonreflexive simulation mechanisms …" (Gallese 2005, p. 117). He refers to his model as the "shared manifold hypothesis" and distinguishes between three levels (Gallese 2001, p. 45):

- The *phenomenological level* is the one responsible for the sense of similarity … that we experience anytime we confront ourselves with other human beings. It could be defined also as the *empathic* level ….
- The *functional level* can be characterized in terms of simulation routines, *as if* processes enabling models of others to be created.
- The *subpersonal level* is instantiated as the result of the activity of a series of mirror matching neural circuits.

As we saw, one possible objection to ST is that since it employs a model that is first person, or at least confined to my own system (a simulation in my own mind or motor system) nothing justifies inferring anything about what must be going on in the other person. As Hutto (in press) suggests, "what is needed is a reliable method of indicating the *other's* perspective on events, not a projection of our own on or into theirs." Does a subpersonal simulation lock us up within our own first-person system?

Defenders of the implicit version of ST have an answer to this. Mirror neurons (and shared representations) are, it is claimed, *neutral* – neither first nor third person – they are activated both for my own action and for observation of the

other's action: Activation of the system simulates the intentional action but does not identify the agent (deVignemont 2004; Gallese 2005; Hurley 2005; Jeannerod and Pacherie 2004). In this case, the subpersonal simulation process, like its explicit cousin, involves a multi-step process. First, we perceive the other's behavior; this is followed immediately by activation of shared representations – in neutral mode (registering "naked intentions"); and this is followed by a determination of the agent, i.e., a specification of who did the action – me or the other person (Jeannerod and Pacherie 2004). This final step is accomplished by what Georgieff and Jeannerod (1998) call the "Who" mechanism.

It will pay to stop and consider Jeannerod and Pacherie's claims about naked intentions. They assume that an articulation at the level of the neural activations between those responsible for (1) registering in the perceiving system the ("naked") intention in an action, and (2) registering the agency for the action, means that there is an articulation in experience between the perception of intention and the experience of agency. "We can be aware of an intention, without by the same token being aware of whose intention it is. … something more than the sole awareness of a naked intention is needed to determine its author" (2004, p. 140). If in fact the brain can process information about intentions without assigning agency to the intentions, is it legitimate to say that our experience is similarly articulated? Jeannerod and Pacherie suggest that it is.

> When the naked intention one is aware of yields an overt action, the extra information needed to establish authorship may be found in the outside world. The question 'Is this intention mine?' would then be answered by answering the question: 'Is this my body performing the corresponding action?' (2004, p. 140).

Phenomenologically (experientially), however, intentions in almost all cases come already clothed in agency. The "who" question hardly ever comes up at the level of experience, because the neural systems have already decided the issue – one way or the other – i.e., even if I'm wrong about who is acting, I am still experiencing or perceiving the intention as already determined in respect to agency. The wonderful thing about the "Who system" is that it is completely neurological and subpersonal – and the results of its activation are hardly ever experientially manifested as "making a decision about who did the action." Rather, the results of its activation are experientially manifested as "X's action" where X is either you or me. Indeed, our direct perception is highly reliable in regard to discriminating between self and nonself. Pathologies and oddly arranged experiments may reveal "who" problems, but in normal ecological behavior it is generally clear whose intention/action it is. As we know from philosophers like Wittgenstein, Shoemaker, and Evans, the identification question – "Someone is intending to pick up the apple, is it me?" – just doesn't come up. There is, in effect, no necessary isomorphism between the phenomenological level and the neuronal level. So if the neuronal processes can be defined as involving a step-wise process, this does not mean that a step-wise process needs to show up in phenomenology.[5]

This brings us to the first set of questions about implicit ST, based on an alternative interpretation of neural resonance. Implicit ST interprets neural resonance as

simulation; but it could easily be interpreted as part of the neuronal processes that underlie intersubjective *perception* rather than simulation. That is, the articulated neuronal processes that include activation of mirror neurons or shared representations may underpin a nonarticulated direct perception of the other person's intentional actions, rather than a distinct mental process of simulating their intentions. This claim requires that we conceive of perception as a temporal phenomenon and as an enactive sensory-motor phenomenon.

First, mirror neurons fire 30–100 ms after appropriate visual stimulation (Gallese, private correspondence). What is, even in neurological terms, a short amount of time between activation of the visual cortex and activation of the premotor cortex, raises the question of where precisely to draw the line between the act of perception and something that would count as a simulation. Even if it is possible to draw a line between activation of the visual cortex and activation of the premotor cortex, this does not mean that this line distinguishes between perception and simulation as a step-wise process.

To be sure, Gallese and the implicit simulationists are not claiming that the step-wise neuronal processes (sensory activation of visual cortex followed by mirror system activation) generate a step-wise conscious process of perception plus simulation. Gallese contends that the simulation stays implicit; that the mirror system activation itself can be read, functionally, as a simulation process. Nonetheless, I think two issues can be raised in this regard.

First, deciding that mirror neurons function as simulations depends on taking a step-wise model that was developed at the explicit, conscious, or personal level, and looking for that step-wise model at the neuronal level. Simulation, according to various versions of ST, involves a step-wise process that begins with perception and ends with some form of understanding. We first see an action that we need to understand; we then simulate it in our own mind or motor system; we then attribute agency for the action, or infer or understand something about the other's experience. But if neuronal processes that send information from sensory cortex to premotor cortex take some time (as much as 100 ms), it is not clear that we should identify two functionally distinct steps rather than a temporally extended and enactive perceptual process. That is, at least in terms of temporal parameters, the fact that at the neurological level, S (sensory processing) is followed by M (activation of mirror neurons) does not mean that one should distinguish between perception and simulation.

If we think of perception as an enactive process (e.g., Noë 2004; Hurley 1998) – as sensory-motor, and not just as sensory input – then it may be more appropriate to think of the resonant processes as part of the structure of the perceptual process when perception is of the action of conspecifics or members of close species. Mirror activation is not the initiation of simulation; it's part of a direct perception of what the other is doing. At the phenomenological level, when I see the other's action or gesture, I *directly perceive* the meaning in the action or gesture. I see the joy or I see the anger, or I see the intention in the face or in the posture or in the gesture or action of the other. I see it. I don't have to simulate it. And I immediately see that

it is *their* action, gesture, emotion, or intention, and it is extremely rare that I would be in a position to confuse it with my own. This perceptual access to the other's intentions may be based on the underlying articulated neuronal processes, but it is not clear at all why we should think of these processes on the simulationist model.

Of course the simulationist can accept the phenomenology ("Yes indeed, that is what *seems* to happen") and still hold to the interpretation that these specific subpersonal processes involve simulation. In fact, however, if we use Gallese's distinctions, this interpretation goes from the functionalist level to the neurological level, but with no clear justification. That is, what happens on the neurological level is simply a complex sequence of neuronal activations. If we look at those processes from a functionalist perspective already framed by ST, then we tend to read those processes as involving simulation. If, in contrast, we look at those processes from a phenomenological level that suggests a direct perception of the other's intentions, then we tend to read those processes as perceptual without simulation. Can the simulationist offer any convincing evidence that the activation of resonance processes is in fact a simulation?

This brings us to a second issue, and more properly an argument against the implicit version of ST. What theorists of implicit simulation (Gallese, Jeannerod, Pacherie) call 'simulation' is not simulation in any genuine sense of the word. Consider, first, two definitions of 'simulation' offered by the OED.

The pretense definition. Simulation is an imitation, in the sense of something *not real* – counterfeit; to simulate means to feign, to pretend. We can find this use of the term in Wittgenstein. "Why can't a dog simulate pain? Is he too honest? Could we teach a dog to simulate pain? Perhaps it is possible to teach him to howl on particular occasions as if he were in pain, even when he is not" (Wittgenstein 1958, Sect. 250).

The instrumentalist definition. Simulation in the sense of a simulator: a *model* (a thing) that we can *use* or do things with so we can understand the real thing.

We can find both senses of the term in the literature of ST. Consider the following characterizations (italics are mine). Simulation means "*using* one's own evaluation and reasoning mechanisms as a model for theirs …" (Dokic and Proust 2002, p. viii). Simulation involves 'pretend states' where

> by pretend state I mean some sort of surrogate state, which is *deliberately adopted* for the sake of the attributor's task … In simulating practical reasoning, the attributor *feeds* pretend desires and beliefs into her own practical reasoning system (Goldman 2002, p. 7).

The pretense is of a precise kind. Bernier (2002) makes this explicit as an essential element found in ST.

> According to ST, a simulator who runs a simulation of a target would *use* the resources of her own decision making mechanism, in an "off-line" mode, and then the mechanism would be fed with the mental states she would have *if she was in the target's situation* (Bernier 2002, p. 34).

For ST, a simulation is not simply a model that we use to understand the other person – theoretical models would suffice if this were all that is required. Even the fact that the model is constituted in our own mechanisms is not sufficient. Rather,

I must use the model 'as if' I were in the other person's situation. As Gallese puts it, "our motor system becomes active *as if* we were executing that very same action that we are observing" (2001, p. 37). Gordon locates this pretense right at the neuronal level: The neurons that respond when I see your intentional action, respond "*as if* I were carrying out the behavior …" (2005, p. 96). Simulation, then clearly needs to meet these two conditions: It is a process that I control in an instrumental way (in the explicit version it is "deliberately adopted"), and it involves pretense (I put myself "as if" in the other person's shoes).

We find these two characteristics in almost every description of simulation in ST. Mental simulation is a cognitive "*ability* or *heuristic* or *methodology*" (Jacob 2002, who cites Gordon for the latter term) – by which I "*engage* in pretense," *put myself* in someone else's shoes, *compare* my experience to their experience, and *predict* their mental state, emotion, or behavior. We *use* ourselves as a model … I *create* in myself some pretend beliefs … and so forth. This is the way simulation is characterized not only by theorists of explicit simulation, but also by theorists of implicit simulation. The pretense condition mentioned by Gallese is accomplished in a simulation considered to be "an interactive *model* of what cannot be known in itself" (2003). At the subpersonal level, the brain in a stepwise fashion is *modeling* the intentional action of others. Gordon (2004, p. 1) suggests that on the "cognitive-scientific" model, "one's own behavior control system is employed as a manipulable model of other such systems. (This is not to say that the "person" who is simulating is the model; rather, only that one's brain can be manipulated to model other persons.)"

Thus, the ubiquitous definition of simulation in ST involves the instrumental use of a first-person model to form third person "as if" or "pretend" mental states. Are either of these conditions met by subpersonal resonance processes? If simulation is characterized as a process that I (or my brain) uses or controls, if this is what simulation is, then it seems clear that what is happening in the implicit processes of motor resonance is not simulation. We, at the personal level, do not *do* anything with the activated brain areas – in fact, we have no access to neuronal activation, and we can't use it as a model. Nor does it make sense to say that at the subpersonal level the brain is *using* a model or methodology, or *comparing* one experience with another, or *creating* pretend states, or that one set of neurons makes use of another set of neurons as a model. These neuronal systems do not take the initiative; they do not activate themselves but are activated by the other person's action. The perception of the other person's action automatically activates in our brain the same areas that are activated when we engage in similar action. The other person *has an effect on us*. The other *elicits* this activation. This is not a simulation, but a perceptual elicitation. It is not us (or our brain) *doing* it, but the other who does this to us.[6]

Furthermore, in subpersonal processes there is no pretense, and this is the case whether we consider neuronal processes as vehicles or in terms of the content that they might represent. As vehicles, neurons either fire or don't fire. They don't pretend to fire. More to the point, however, what these neurons represent or register cannot be pretense in the way required for ST. They do not fire "as if" I were

you. As we saw, the mirror system is neutral with respect to the agent; there is no first- or third-person specification involved. In that case, they do not register my intentions as pretending to be your intentions; there is no "as if" possible in a neutral system like this.

Goldman and Sripada (2005), acknowledging the discrepancy between the ST definition of simulation and the working of subpersonal mirror processes, propose a minimal necessary condition for simulation: "Applied to mindreading, a minimally necessary condition [for simulation] is that the state ascribed to the target is ascribed as a result of the attributor's instantiating, undergoing, or experiencing, that very state. In the case of successful simulation, the experienced state matches that of the target" (p. 208). If this is a necessary condition, it cannot be a sufficient one, because without something further, it is not clear what would motivate me to ascribe the state that I was undergoing to someone else. Furthermore, if this were as automatic as mirror neurons firing, then it would seem that we would not be able to attribute a state different from our own to someone else. But we do this all the time. I can understand that the person in front of me is enthusiastically and gleefully reaching to pick up a slug at the same time that I am experiencing revulsion and disgust about that very possibility. It seems straightforward to say that I see what she is doing and that I see she is doing it with enthusiasm and glee, but that my own feelings are quite different. Furthermore, consider the difficulties involved if we were interacting with more than one other person. Is it possible to simulate the neural/mental/emotional states of two other people at the same time if in fact our simulations must be such that we instantiate, undergo, or experience, those two (possibly very different) states?[7]

It is therefore not clear why we should think of the activation of resonance systems as a simulation process of the sort required by ST. This is not to deny that there are resonance processes at work in our perception of the other person. Moreover, the nature of the resonance processes involved in such encounters makes our perception of other conspecifics different from our perception of objects and instruments. But it doesn't make social cognition the result of an implicit simulation.

4.4. CONCLUSION

I have brought a number of arguments against various versions of ST. What I've called the simple phenomenological argument can challenge both explicit and radical versions, especially if we take these versions to describe our primary, pervasive, or default mode of understanding others. In such cases, ST should be able to call on phenomenological evidence to verify the simulation model. In fact, phenomenological reflection on our ordinary experience speaks against ST. In the same way that in our everyday encounters we find a scarcity of theoretical musing about others, we find a scarcity of simulation routines at work. One could argue, of course, that phenomenology is not always correct, even in regard to the most explicit experiences. But that argument would have to be extended too far if the

claim were to be made that in fact we use simulation routines all the time, but phenomenology constantly and consistently misses this fact.

In addition to what I've called the simple phenomenological argument, phenomenologists and others, like Ryle, have pointed to logical and conceptual problems with ST. One of these objections challenges all versions of ST, even though it was originally brought against the earlier theory based on inference from analogy. Given the large diversity of motives, beliefs, desires, and behaviors in the world, it is not clear how a simulation process based on my own relatively narrow experience (or relatively unique circumstance) can give me a reliable sense of what's going on in the other person's mind, or in their behavior.

Implicit ST may be in a better position to answer this particular charge. Perhaps the common tendency to anthropomorphize (which isn't necessarily bad if the other we are concerned about is another human), or to use the intentional stance, is based on just those motor resonance systems that work on a basic interbodily level. Moreover, the test of ST shouldn't be framed in terms of reliability. After all, we often misunderstand others, and whatever mechanism accounts for intersubjectivity, it will necessarily have some degree of unreliability built in.

The problem with implicit ST is that it is only one, and not necessarily the best interpretation of the significance of motor resonance systems. I've argued that implicit resonance processes are not simulations in any sense that is useful for ST. Furthermore, if implicit ST does give an account of our primary and pervasive ability to understand others, it would count as an argument against explicit ST, since explicit simulations would be redundant in this case. Likewise, however, if our default mode of understanding others were based on explicit simulation, as Goldman contends, then the claims of implicit ST about the adequacy of motor resonance processes would be wrong. Goldman's view of implicit motor resonance processes is that they do not constitute simulations of a sort that would be sufficient to do the full job, but do generate some background information that is useful to initiate the explicit simulation process. I've argued, in contrast to both explicit and implicit ST, that implicit motor resonance processes are important enactive processes that contribute to the constitution of the perceptual access that we have to the intentions of others.

I do not claim that we get a full account of human intersubjectivity in the idea that we have *perceptual* access to the intentions of others. Perceptual access to the other person's bodily movements, gestures, facial expressions, and so forth does give us a sense of what is going on with them, what they mean and what they feel, and this, together with our interactions with them in pragmatic and social contexts, gives us a relatively reliable, but still minimal understanding of them. There is much more to say about the role of language and narrative competency in a fuller account of intersubjectivity (see Hutto, 2003, 2004, forthcoming; Gallagher 2006; Gallagher and Hutto, in press). Even in that larger story, however, the theory of mind approaches that emphasize either simulation, or the role of folk psychology as background theory, have a minimal role to play in our normal and everyday interactions.

NOTES

[1] Hurley (2005) points to this as one difference between TT and ST. "When I use practical reason off-line in mind-reading, I don't formulate normative laws from which I make inferences; rather, I activate my own normative and deliberative dispositions."

[2] Infants are indeed capable of understanding the intentions of others. Meltzoff (1995) shows that children at 18 months of age are capable of recognizing and completing another person's failed intention. The experimenter pretends to have great difficulty accomplishing a certain task with a toy and presents an incomplete action to the child. The child, who takes the toy and completes the task with little effort, demonstrates that she understands what the experimenter desired to do. Even earlier, infants seem capable of parsing intentions (see Baird and Baldwin 2001; Baldwin and Baird 2001). Whether some animals are capable of understanding intentions is still a debated issue.

[3] Third-person mentalization signifies simply that one person is trying to understand another person, rather than trying to understand herself (which would be first-person). Some theory-of-minders contend that we use simulation (or theory) to understand our own minds.

[4] The implicit version of ST actually counts as an argument against the explicit version of ST. That is, if our ability to understand others is in fact mediated by an implicit and automatic simulation process, then we have little need for the more explicit version. Indeed, to the extent that an implicit ST would explain the phenomenological scarcity of explicit simulation, it would support the simple phenomenological argument against explicit simulation. Along this line Gallese (2005, p. 102) states: "Whenever we face situations in which exposure to others' behavior require a response by us, be it active or simply attentive, we seldom engage ourselves in an explicit, deliberate interpretive act. Our understanding of a situation most of the time is immediate, automatic, and almost reflex like."

[5] On the question about isomorphism between subpersonal and personal levels, see Gallagher (1997), Hurley (2005) and Varela (1996).

[6] It may seem contradictory to claim in the previous paragraphs that perception is enactive, or as Noë says, "perception is action," and now to claim that the activation of the resonance system is the result of a passive elicitation, so that the motor aspect of perception does not involve our action, but is a case of us being affected by the other. I think that a fuller account of enactive perception has to be able to accommodate this passive, affective aspect of perception (see Gallagher 2005).

[7] Seeing a person reaching for a slug is a good example with which to make one more clarification. Neuroscientists sometimes use the term 'simulation' to refer to certain motor control processes for action planning. Efference copy sent through forward control mechanisms, for example, is said to constitute a simulation of an intended movement in order to compare it with an ongoing movement to predict its success. The brain runs this simulation to make fast non-conscious corrections to keep the action on track. This use of the term is entirely independent of ST for social cognition, and the objections I am raising here do not apply to it. Some theorists, however, have appealed to these motor emulation processes as possible mechanisms involved in the simulation of another's action (e.g., Gallese 2001; Hurley 2005, pp. 181–188; see Iacoboni, cited in Millikan 2005, p. 188n2). The problem, again, is that if indeed the subpersonal emulation is neutral in regard to who's action is at stake, the pretense condition is not met; there is no "*as if* it were I" involved, and in that regard it fails to be the kind of simulation required by ST. Then it can be only a representation of an intentional action in my motor system, but not a representation of my own motor action as if it were the other's.

REFERENCES

Baird JA, Baldwin DA (2001) Making sense of human behavior: action parsing and intentional inference. In: Malle BF, Moses LJ, Baldwin DA (ed) Intentions and intentionality: foundations of social cognition. MIT Press, Cambridge, MA, pp 193–206.

Baldwin DA, Baird JA (2001) Discerning intentions in dynamic human action. Trends Cogn Sci 5(4):171–78.

Bernier P (2002) From simulation to theory. In: Dokic J, Proust J (eds) Simulation and knowledge of action. Benjamins, Amsterdam, pp 33–48.

Currie G, Ravenscroft I (2002) Recreative minds. Oxford University Press, Oxford.
deVignemont F (2004) The co-consciousness hypothesis. Phenomenol Cogn Sci 3(1):97–114.
Dokic J, Proust J (2002) Introduction. In: Dokic J, Proust J (eds) Simulation and knowledge of action. Benjamins, Amsterdam, pp vii–xxi.
Evans G (1982) The varieties of reference. In: McDowell J (ed). Oxford University Press, Oxford.
Fadiga L et al. (1995) Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol 73:2608–2611.
Gallagher S (1997) Mutual enlightenment: recent phenomenology in cognitive science. J Conscious Stud 4(3):195–214.
Gallagher S (2001) The practice of mind: theory, simulation, or interaction? J Conscious Stud 8(5–7):83–107.
Gallagher S (2004) Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. *Philos Psychiatry Psychol* 11(3):199–217.
Gallagher S (2005) Book review: a new movement in perception: review of Alva Noë's action in perception. Times literary supplement (London).
Gallagher S (2006) The narrative alternative to theory of mind In: Menary R (ed) Consciousness and Emotion: special issue on radical enactivism.
Gallagher S, Hutto D (in press) Primary interaction and narrative practice: from empathic resonance to folk psychology. In: Zlatev, Racine, Sinha, Itkonen (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam.
Gallese V (2001) The 'shared manifold' hypothesis: from mirror neurons to empathy. J Conscious Stud 8:33–50.
Gallese V (2003) The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36:171–180.
Gallese (2005) 'Being like me': self-other identity, mirror neurons and empathy. In: Hurley S, Chater N. (eds) Perspectives on imitation I. MIT Press, Cambridge, MA, pp 101–118.
Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. Trends Cogn Sci 2:493–501.
Georgieff N, Jeannerod M (1998) Beyond consciousness of external events: a 'Who' system for consciousness of action and self-consciousness. Conscious Cogn 7:465–477.
Goldman AI (2002) Simulation theory and mental concepts. In: Dokic J, Proust J (eds) Simulation and knowledge of action. Benjamins, Amsterdam, pp 1–19.
Goldman A (2005a) Mirror systems, social understanding and social cognition. Interdisciplines (http://www.interdisciplines.org/mirror/papers/3).
Goldman A (2005b) Imitation, mind reading, and simulation. In: Hurley S, Chater N (eds) Perspectives on imitation II. MIT Press, Cambridge, MA, pp 80–81.
Goldman AI, Sripada CS (2005) simulationist models of face-based emotion recognition. Cognition 94:193–213.
Gordon RM (1995a) Folk psychology as simulation. In: Davis M, Stone T (eds) Folk psychology. Blackwell, Oxford, pp 60–73.
Gordon RM (1995b) Developing commonsense psychology: experimental data and philosophical data. APA eastern division symposium on children's theory of mind, December 27, 1995 (http://www.umsl.edu/~philo/Mind_Seminar/New%20Pages/papers/Gordon/apakids9.htm).
Gordon RM (2004) Folk psychology as mental simulation. In: Zalta N (ed) The Stanford encyclopedia of philosophy (URL=http://plato.stanford.edu/archives/fall2004/entries/folkpsych-simulation/).
Gordon RM (2005) Intentional agents like myself. In: Hurley S, Chater N (eds) Perspectives on imitation I. MIT Press, Cambridge, MA, pp 95–106.
Gordon R, Cruz J (2003) Simulation theory. Encyclopedia of cognitive science. The Nature Publishing Group Macmillan Reference Ltd, London.
Grèzes J, Decety J (2001) Functional anatomy of execution, mental simulation, and verb generation of actions: a meta-analysis. Hum Brain Mapp 12:1–19.
Hurley SL (1998) Consciousness in action. Harvard University Press, Cambridge, MA.

Hurley SL (2005) Active perception and perceiving action: the shared circuits model. In: Gendler T, Hawthorne J (eds) Perceptual experience. Oxford University Press, New York.
Hutto D (2003) Folk-psychological narratives and the case of autism. Philos Pap 33(3):345–361.
Hutto D (2004) The limits of spectatorial folk-psychology. Mind Lang 19:548–573.
Hutto D (in press) Unprincipled engagements: emotional experience, expression and response. In: Menary R (ed) Consciousness and emotion. Book series, vol. 7. Special issue on radical enactivism.
Hutto D (forthcoming) Folk psychological narratives. MIT Press, Cambridge, MA.
Jacob P (2002) The scope and limits of mental simulation. In: Dokic J, Proust J (eds) Simulation and knowledge of action. Benjamins, Amsterdam, pp 87–109.
Jeannerod M, Pacherie E (2004) Agency, simulation, and self-identification. Mind Lang 19(2):113–146.
Meltzoff AN (1995) Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Dev Psychol 31:838–850.
Millikan RG (2005) Some reflections on the theory theory–simulation theory debate. In: Hurley S, Chater N (eds) Perspectives on imitation II. MIT Press, Cambridge, MA, pp 182–188.
Mitchell JP (2005) The false dichotomy between simulation and theory–theory: the argument's error. Trends Cogn Sci 9(8):363.
Nichols S, Stich S (2003) Mindreading. Oxford University Press, Oxford.
Noë A (2004) Action in perception. MIT Press, Cambridge, MA.
Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Cogn Brain Res 3:131–141.
Rizzolatti G, Fogassi L, Gallese, V (2000) Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions. In: Gazzaniga MS (ed) The new cognitive neurosciences. MIT Press, Cambridge, MA, pp 539–552.
Ryle G (1949) The concept of mind. Barnes and Noble, New York.
Scheler M (1954) The nature of sympathy. Transl. by Heath P. Routledge and Kegan Paul, London. Original: Wesen und Formen der Sympathie. Verlag Friedrich Cohen, Bonn, 1923.
Varela FJ (1996) Neurophenomenology: a methodological remedy for the hard problem. J Conscious Stud 3(4):330–349.
Wittgenstein L (1958) Philosophical investigations. Transl. by Anscombe GEM. 3rd edn. Macmillan, New York.

BEATA STAWARSKA

5. PERSONS, PRONOUNS, AND PERSPECTIVES

Folk psychology has emerged as the dominant approach to social cognition within the analytic philosophy of mind. Its basic tenet is conveyed by the allegedly commonsense conviction that understanding other mindful beings consists in the ability to attribute intentional states, notably beliefs and desires, in order to predict and explain their behavior. Folk psychology defendants may differ as to what strategies are employed in the attribution of intentional states. While the proponents of the so-called *theory* theory of mind argue that the strategies recruited in the understanding of minds consist of detached theoretical procedures, the defendants of the so-called *simulation* theory of mind point to more engaged, simulational routines, with affective content and practical significance. A recent trend in the discussion suggests that a combination of both theory and simulation may be involved in social cognition, and so rather than oppose them, we should include both the theoretical and practical strategies in a comprehensive story of how humans make sense of minds. This recent rapprochement between theory and simulation theorists is indicative of the fact that, despite their overt differences as to the mechanisms employed in social cognition, they share a deep-seated conviction regarding the character of social cognition. On both accounts, social cognition is a process of *mentalizing*, i.e., attributing hidden mental states to other people (in order to predict and explain their manifest behavior). Both accounts thus operate with a split between the outer physical datum and the inner mental content. Furthermore, by focusing on prediction and explanation of behavior, both accounts tacitly assume that social life involves *an observational stance* towards others' publicly available mimicry, gestures, and speech, in view of extracting their mental underpinnings, rather than the stance of direct interpersonal interaction based on reciprocal engagement between social partners. As Hutto (2004) recently argued, received thinking about folk psychology is predicated upon the assumption that our initial stance with regard to others is essentially estranged and that we follow what Bogdan (1997) termed the "spectatorial view of interpretation" by regarding others as remote objects of study rather than as social partners. Consequently, on this understanding sociality privileges *a third-person approach* towards one's fellow beings, *about* whom one needs to theorize or whom one needs to model by means of simulational routines, at the exclusion of the *second-person* approach, where the interaction is a direct source of mutual understanding.

These two key assumptions about the nature of social cognition (it is a mentalizing process, adopting a third person or observational stance) point to serious limitations of folk psychology. As Gallagher (2001) argued, folk psychology offers an overly intellectualized picture of social interaction. Furthermore, it is unclear whether we

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 79–99.

ordinarily resort to the strategies of explaining and predicting behavior – unless we are at a loss to make sense of what the other is up to. In that case, however, it is unclear how folk psychological methods could positively advance our understanding of the other's puzzling behavior (see Hutto 2004 for a discussion of this point). Rather than rely on the intellectually heavy methods of mind-reading which operate at a theoretical remove from others, we should rather consider, following Gallagher (2001) that social life takes the form of "embodied practice" where we interact with others through gesture, facial expression, eye contact and gaze tracking, both in infancy and throughout a typically developing adult life. These primary noninferential forms of engaging others are foundational to social cognition in that they are not only primary in developmental terms, but also "primary across all face-to-face intersubjective experiences" (Gallagher 2001, p. 103). Contra the claims made by the friends of folk psychology, it is embodied nonconceptual second-person interaction with others that provides the key to any experience-based theory of social life.

It has recently become a trend among the philosophers whose research bridges the purported divide between the continental and the analytic philosophical traditions to turn to classical phenomenology in the search for approaches to sociality that pay heed to the lived experience of interpersonal interaction and to offer phenomenological critiques of the folk psychological models of social life. Unquestionably, the phenomenological tradition offers multiple rich and instructive contributions to the question of sociality. However, classical phenomenology has been informed by the egological tradition (discussed below), and, as such, it is itself subject to the critique advanced against folk psychology. If my argument is sound, folk psychology and classical phenomenology share an excessive reliance on the mentalizing third-person stance towards others. I expand on this critical thought in Sect. 5.1 of this chapter. I then proceed in Sects 5.2–5.4 to promote an alternative to both the folk psychological and classical phenomenological approaches, which underscores the primacy of second-person interaction – or a dialogic model of I-you connectedness. I accomplish this by combining contributions from the following traditions of inquiry: linguistics and phenomenology (Sect. 5.2), developmental psychology (Sect. 5.3), and, in brief, the philosophy of dialogue and ecological psychology (Sect. 5.4). The criticism advanced in part 1 is not indispensable to appreciating the discussion of the primacy of second-person relations and their irreducibility to third-person relations (Sects 5.2–5.4). Rather, it provides a backdrop against which the positive contributions of this chapter are made.

5.1. THE EGOLOGICAL TRADITION AND CLASSICAL PHENOMENOLOGY

In examining the classical phenomenological thought, I will focus on the extent to which the phenomenology of consciousness is informed by the *egological tradition*, and how this constrains phenomenological accounts of interpersonal relations. My thinking here echoes the statements made by Gurwitsch (1979) about the "traditional

theories" of social relations. Gurwitsch grouped Husserl (of the *Cartesian Meditations*) together with Kant, Locke, and Descartes, under the heading of the traditional theory that pervades the phenomenology of consciousness. The pronouncements of the four philosophers about the social world were, Gurwitsch maintained, derivative of the overriding interest in securing the foundations of philosophic knowledge. The "traditional" account is therefore not a purely descriptive one but driven by an underlying epistemological agenda, which shapes the resulting theory of intersubjectivity in determinate ways, notably by centering the field of inquiry in individual consciousness accessed in the first-person mode. It does not therefore come as a surprise that, as Gurwitsch has it, one of the defining components of the traditional approach is that conscious mental processes have an irreducible relation or "appertinence" to the ego; the ego is a necessary "index" born by mental states that enables their identification as my own (pp. 1–2). Gurwitsch believed that such an egological construal of the mind places the phenomenology of consciousness in sharp contrast to the data of lived human experience, notably the nontheoretical conviction that we are in immediate perceptual presence of other persons whose mental life is apparent in and through their manifest behavior (p. 3). Gurwitsch was skeptical therefore about the ability of the phenomenology of consciousness to ever produce an adequate rendering of "human encounters in the social world." Within the egological construal of the mind, the passage from the ego to other human beings becomes barred, or in Gurwitsch's own words, "mental processes appertinent to 'We' " [*Wir-Erlebnisse*] become unintelligible" (p. 28).

Following Gurwitsch's lead, I propose to deepen the analysis of how the egological framework impacts phenomenological discussions of sociality. I will view egology, however, through an explicitly linguistic lens. The term *ego* has become so entrenched in post-Cartesian philosophy that a reminder of its original function in ordinary language may be in order. Before it was imported into the *lingua philosophica*, prefaced with a definite article, and handled as if it were a noun, the *ego* served in its native Latin as the first-person pronoun, as in the classic statement *ego cogito* or *I think*. This import of the first-person pronoun into technical jargon and its concurrent nominalization have, to borrow an apt phrase used in a slightly different context by the French linguist Emile Benveniste (1971, p. 62), "set [it] up … as an objectifiable notion which philosophical thought could handle, analyze, and define just as any other concept."[1] However, as will be discussed in part 2 in reference primarily to Benveniste, the pronoun *I* (or *ego*) does not behave just like any other concept within ordinary language, notably it does not behave like a noun designating an objective entity, whether it be one's own self – as in "my ego" – or the other person – as in "the alter ego." Forcing the pronoun *I* (or *ego*) to behave like a noun is a theoretical violation, which motivates a miscontrual of the relations between self and others in egological philosophy. By returning the pronoun *I* (or *ego*) to its ordinary function in the natural language, it will be possible to both challenge the egological tradition *and* to point to a philosophic framework better adapted to the reality of the social world.

Before examining the ego and other pronouns as explicitly linguistic categories, I need to expand upon the status of the ego in Husserl's classical phenomenological formulations, with brief references to Descartes as well as Merleau-Ponty and Sartre. My analysis will be confined to a number of systematic points – I leave it to Husserl scholars to provide a textually richer interpretation of the founding father's entire *corpus* as well as to determine the possible extent of his departure from the egological framework in the texts postdating Gurwitsch's criticism of the phenomenology of consciousness. Restricted in the aforementioned manner, my analysis will challenge classical phenomenology's purported ability to critique and transcend the folk psychological model of social cognition. It may be that phenomenology needs to revise its own theoretical base if it is to provide viable alternatives to the models of sociality currently in use in the cognitive sciences. This is definitely not to exclude the possibility of phenomenological investigations unconstrained by the egocentric model but to expose how such constraints harm rather than help a phenomenological investigation of sociality.

As is well known to all philosophical children, the success of Descartes' project carried out in *The Meditations* is measured by his locating absolutely certain knowledge within the domain of the conscious mind. Having subjected to doubt common existential beliefs regarding the world as well as the validity of scientific concepts, it remains necessarily true that the doubting/thinking self – subsumed under the Latin pronoun *ego* – must necessarily exist. In other words, the ego stands as the irreducible residuum of the suspension of belief in extra-egological data, whether they are mundane things, animate bodies, or scientific concepts. The ontological status of this hypothetical survivor was for Descartes unquestionably that of an immaterial substance, a thinking thing. It is this theoretical move, which Husserl later found to be problematic. As he puts it, "in the foundation-laying reflections of the *Meditations* – those in which the epoche and its ego are introduced – a break of consistency occurs when this ego is identified with the pure soul" (1970, pp. 79/80).

To remedy this shortcoming, the Husserl of the *Cartesian Meditations* re-enacted "the necessary regress to the ego" (1991, p. 6), in a way that capitalizes on Descartes' ground-breaking methodology while lifting the objectivist bias of his philosophy. Having envisaged "the possible nonbeing of the world," Husserl uncovered a field that is "absolutely subjective" (1991, p. 30), and thus safeguarded the ego's transcendental purity by consistently bracketing the contents belonging to the natural attitude from the philosophical inquiry. He was thus able to lay the foundations of the new science of "pure egology" (Husserl 1991). This purging of the egological field from mundane residua distantiates Husserl, in his own view, from his famous predecessor and trailblazer; both Husserl and Descartes share, however, the epistemological objective of securing apodictic knowledge within the domain of egological consciousness, and their thinking about social relations will be consistently filtered through this epistemological concern.

It remains to inquire into how we are to conceive of the transcendental ego and the science based upon it. For one, the transcendental ego is not to be equated with the empirical human being. However, that does not imply that it stands for a supraindividual, universal subject. Nor is it an empty subjective pole – Husserl attributes facticity to the ego (2002, p. 313). Finally, and most importantly, the transcendental ego is also my own ego, belonging to the so-called sphere of *ownness*. That does not exclude other egos from the field of phenomenological investigation, insofar as phenomenology's task is to provide a viable theory of transcendental intersubjectivity. It does, however, mean that my ego is both an initial and a unique object of study. My ego is "the one and only absolute ego" (1970, p. 69). Unsurprisingly then, my transcendental ego is also absolutely solitary. As Fink put it, enacting the phenomenological reduction places me "in the monstrous solitude of transcendental existence [*Existenz*] as ego" (1995, p. 99).

Insofar as the transcendental ego is absolved from the mundane and human domains, it does not come as a surprise that Husserl regards it as absolved also from "the whole distinction and ordering of personal pronouns" (1970, p. 184), as they function in a natural language. Put differently, the word of the first-person pronoun (Latin *ego*, German *Ich*, English *I*…) is preserved in the transcendental domain, but the whole diacritical network connecting it in the ordinary language context to other words, notably to other personal pronouns, is put out of play. As Husserl (1970) puts it, "I am not *an* ego, who still has his *you*, his *we*, his total community of cosubjects in natural validity." On the transcendental plane, I am not *an* ego amongst others. For even though the task of phenomenology is to ground a transcendental theory of intersubjectivity, still this objective must not have us "leap over the primal "I," the ego of my epoche, which can never lose its uniqueness and personal indeclinability" (1970, p. 185).

Before discussing this "indeclinable I" in more detail, consider that natural or ordinary language, considered as a mundane existent, is subject to the phenomenological reduction. This raises an important question, which occupied Fink in section 10 of the *Sixth Cartesian Meditation* (the entire text was approved by Husserl): What language, if any, would one then speak (read, write) in the transcendental domain? Fink claimed that even though the entry into the transcendental domain requires the bracketing of the natural language, that does not mean that a new transcendental language should be devised in its place. It is rather the question of the natural language conveying different meanings than it conveyed hitherto. No longer an end product of the constituting activity, belonging to the realm of the existent world, but rather the medium enlisted to capture that very constituting activity of consciousness, language is required to perform a task exceeding its available resources, it is forced, so to say, to outperform itself. Transcendental statements are therefore necessarily plagued with equivocation – they never mean just what they say. Hence the "immanent conflict and contradiction in every transcendental predication" (p. 98) and "the constant rebellion against the constraint imposed upon [the transcendental sense] by the formulation in natural words and sentences" (p. 89). In fact, Fink contends that the phenomenologist who adopts

the transcendental attitude finds herself at first lacking language, and goes on to wonder whether "this experiential life [could not] remain then *forever* without language?" (p. 98). To which he responds, in two discordant statements, that "there is *no reason* and *no compulsion* for predicative outward expression lying in phenomenologically theorizing experience as such. And yet predicative outward expression is in a definite sense *transcendentally necessary*" (pp. 98/99). This quasi-simultaneous denial and affirmation of the necessity to verbalize transcendental insights contains a blatant contradiction, which may be softened by Fink's invoking an inherent "tendency towards the universal" as the motive for linguistic expression of transcendental meanings (p. 99). He further points to "the communicative tendency of all philosophizing" and stresses that "phenomenologizing [must] be expressed outwardly" (p. 100). Be it as it may, the thrust of Fink's argumentation is that the natural language serves as a means of transport for the insights previously obtained by the phenomenological onlooker without any linguistic intervention.

It is justified to wonder, however, whether a phenomenologist does not necessarily enter the transcendental domain with a linguistic baggage, handed down by the social and cultural tradition she is a part of, which shapes and determines her transcendental insights. The point in question was made by Strawson (1966). Targeting Descartes' rather than Husserl's conception of the conscious mind, Strawson challenged the possibility of positing an individuated consciousness accessed in the first-person mode independently of pronominal categories. Strawson argued that it is the first-person pronoun tacitly employed by Descartes for personal reference that enables him to theorize a single mind at all. In other words, Descartes continues to rely on pronominal categories as he uncovers the indubitable domain of conscious thought with the classic statement 'I think, I am,' even though pronouns in particular and language in general must be allocated to the mundane domain whose existence is subjected to doubt. Descartes' project is therefore inherently absurd: The philosopher withdraws the first-person pronoun from the ordinary language game and recasts it within the domain of pure thought, where it is forced to perform a task of designating an individual enduring consciousness in the first-person mode – a task that it can de facto play only within language, not outside of it.[2] The whole conception of a pure consciousness accessed in isolation from the socio-linguistic community by means of introspective insight is therefore founded on a contradiction; it needs to be replaced, Strawson (1958) contends, with the thesis of the primitiveness of the person, which designates neither the mind nor the body, but is a notion more primary than either of these two terms (see Hobson and Ratcliffe in this volume).

Turning now to Husserl's "critical reinterpretation and correction of the Cartesian conception" (1970, p. 184), could a similar charge be raised to the founder of phenomenology that his conception of the transcendental consciousness is informed by linguistic resources at the very moment he deems them external to the realm of egology? To follow Strawson's line of thought, the ego is not a concept of pure thought revealed to the phenomenologist who adopts the posture of the epoche,

but an element of her active vocabulary, specifically the philosophical jargon which facilitates the use of the first-person pronoun as if it were a noun, thus turning it into an objectifiable notion. This objectivist stance towards the ego is not based solely on insights gained by transcendental reflection but informed by a longer tradition termed by Gurwitsch the phenomenology of consciousness. Not only is the domain of consciousness shaped by ordinary language, as Strawson argued in reference to Descartes, but also by the specific transformations that took place in the philosophical idiom post Descartes. By attributing "personal indeclinability" to the ego, Husserl introduced a notable additional transformation into the philosophical idiom; his transcendental inquiry continues therefore to be directly informed by language, even though the language gets redesigned to fit the interests of the phenomenology of consciousness and is strikingly at odds with common speech.

Consider the puzzling notion of "personal indeclinability" of the transcendental *I* or ego in more detail. As indeclinable, the first-person pronoun becomes absolved from the categories of *number*, *case*, and *gender*. The question of the pronoun's absolution from gender does not affect the qualities it ordinarily bears: The first, unlike the third, person pronoun is gender neutral in all languages under discussion. However, the absolution from the categories of number and case do significantly affect the status of the pronoun. The ordinary I can be put into the plural *We*, but the transcendental *I*, in Husserl's own words, cannot. As indeclinable, the transcendental *I* is then construed as solitary in an inescapable way. It will not suffice to add that this inescapably solitary *I* may serve as a medium through which the plural *we* and other pronouns are grasped nonetheless, primarily by analogy. As Merleau-Ponty, who nota bene adheres to Husserl's notion of the indeclinable *I* in the chapter "Other Selves and the Human World" from the *Phenomenology of Perception* notes, the question then still remains "how can the word 'I' be put into the plural, how can a general idea of the *I* be formed, how can I speak of an *I* other than my own, how can I know that there are other *I*'s, how can consciousness which, by its nature, and as self-knowledge, is in the mode of the *I*, be grasped in the mode of Thou, and through this, in the world of the 'One'?" (1994, p. 348). Furthermore, even if a plurality of *I*s were to be posited, with each defined as indeclinable and serving as an index of supposedly inalienable subjectivity, the social world would be predicated upon a thesis of a multiple solipsism, which is absurd (1994, p. 359). Emancipated from the rules of declination and, by extension, released from the constraints of ordinary grammar, the indeclinable *I* is concurrently removed from the ordinary sociolinguistic framework and so it does not provide the basis for establishing a viable theory of social *relations*, even though it may allow for positing a series of solitary selves.

Moving on to the grammatical category of case, the ordinary *I* can be put in genitive, dative, accusative, and the remaining cases, but not the transcendental *I*. For example, the transcendental *I* does not decline into a *me*, since such declination would presuppose an interpersonal framework within which the perspectives of others as well as my own are in play. The multi-perspectival structure of ordinary

personal pronouns is therefore suspended in the transcendental realm – the only operational perspective in the transcendental field is my own. This uni-perspectival indeclinable *I* can hardly serve as the foundation for a theory of interpersonal relations, where the co-existing perspectives of self and other would have equal primacy.

Consider that even though the grammatical rules of declination may be 'bracketed' in the case of the transcendental ego, this bracketing does not emancipate the ego from grammatical constraints of case and number altogether, but rather redefines them in a way which is at odds with ordinary grammar. The transcendental ego becomes confined to a *single* number and case, namely *the singular nominative*. And so, has it become subjected to, so to speak, a negative variant of ordinary grammar, which prohibits conjointly the use of the plural number and other-than-nominative cases. Such a prohibitive gesture alters – but does not eradicate – grammatical rules. Instead, it imposes a set of 'transcendental' grammatical regulations, which confine the range of applicable cases and numbers to just *one*.

The philosophical implications of these grammatical revisions are that the ego has become construed as an isolated entity, confined to an islet of incommunicability and inescapable solitude, reflecting the supposedly ineffable core of consciousness. The ego has been forced into an egocentric stance. Furthermore, the transcendental ego embodies the two characteristics previously found in the approach to social cognition within folk psychology: (1) it is *mentalistic* and (2) it engages others in *third-person terms*. Contrary to its empirical equivalent, the transcendental ego is construed as a mute or ineffable pronoun, a content of the pure mind, accessible exclusively in the first-person mode. Unlike the first-person pronoun which has general validity, the transcendental ego gets confined to the sphere of ownness, to be accessed exclusively from within.

The other consequence of translating the first-person pronoun from ordinary language into *transcendentalese* is the substitution of a third-person type of relation to others for the second-person type. Recall that, following Husserl, the transcendental ego does not have a Thou or cannot engage the other in a second-person form of direct address. However, in direct linguistic interaction, i.e., in dialogue, the other person is typically addressed in the second-person mode. I say: "What do *you* think about the results of the last presidential election?" Following the prerogatives of transcendental grammar, however, it would be more appropriate to cast this question in third-person terms. Insofar as the transcendental ego is incapable of commanding pronouns other than the first singular, even though it may be handled as a noun and attributed both to self and to others, the above question would thus need to be phrased as "What does the *other I*/*alter ego* think of the results of the last presidential election?" Note that such phrasing has the consequence of making the other person appear primarily as a third party spoken *of* rather than the interlocutor spoken *to*. We will need, however, some additional input from linguistics to gauge the exact importance of this shift from the usual direct form of address in second-person terms to the third-person approach.

5.2. LINGUISTIC CONTRIBUTIONS

Personal pronouns have been famously termed 'shifters' (Jakobson 1973), insofar as their referents are not assigned in a fixed fashion but shift according to the evolving conversational context. Who exactly *I*, *you*, *she*, *he*, *we*, and *they* are is not defined in a stable manner but depends on who speaks, who is addressed, and who is spoken about at a given moment. Personal pronouns are indissociable from the shifting speaker, addressee, and third party roles assumed in the context of dialogue. Despite their shared 'shifty' character, the pronouns do not, however, form a homogenous category, which could be captured in a single definition. As Benveniste (1971), who provides the central point of reference for this section, argues, the traditional nomenclature that aligns the pronouns in a paradigm of conjugation with three terms (*I*, *you*, *s/he*) creates the impression of symmetry between the three persons, but this symmetry is merely formal (p. 221). To be sure, a quick peek into the dictionary suffices to confirm the impression of homogeneity; the dictionary definition, in line with the etymology of the term (from Latin: *pro-nomen*), states that pronoun is a word serving to replace a noun phrase. For example, "the next-door neighbor," "my best friend," "Alice and Elisabeth" can be replaced by *she*, *he*, and *they*, respectively. However, unlike the third-person pronouns, first- and second-person pronouns do not serve to replace a noun phrase and so do not fit the standard definition. "I" does not replace my proper name or an applicable common noun (woman, professor, philosopher, etc.), and neither does "you." They do not designate a previously established referent and so, unlike third-person pronouns, have no representational function.

Hence the need to distinguish the first- and second-person pronouns from the third person one. Benveniste claims in fact that only the former two can justifiably be termed *personal*, while the third designates a nonpersonal entity, which could easily refer to a thing. He follows the definitions used by Arab grammarians, for whom the first person designates 'the one who speaks,' the second 'the one who is addressed,' while the third 'the one who is absent' (p. 197). The first and second persons designate the speaker and addressee, respectively, but not as real objective entities external to the act of speaking. The first person is inextricably tied to what Benveniste calls the instance of discourse, i.e., the actual activity of speaking to one another. This actuality of interpersonal communication is what distinguishes discourse from language, the latter being a system of signs and syntactical rules which could be exemplified by an 'impersonal' scientific text where neither *I* nor *you* need ever be mentioned. In contrast, in discourse the first-person pronoun is necessarily taken up by the speaker as she addresses another by means of the second-person pronoun. Both pronouns refer exclusively to the speaker and addressee roles as they are dialogically deployed. That is why Benveniste claims that *I* and *you* refer solely to the act of individual discourse, where *I* signifies "the person who is uttering the present instance of the discourse containing *I*," while you stands for the "individual spoken to in the present instance of discourse containing the linguistic instance *you*" (p. 218). The referents of *I* and *you* exist therefore exclusively within the dialogic space of direct address, and never outside

it. That is why Benveniste views persons as linguistically constituted, specifically as constituted by the condition of dialogue (p. 224).

Let us examine this dialogical constitution of person in some more detail, in order to prepare the ground for spelling out the distinction between first- and second-, and third-person pronouns. For Benveniste the person arises solely via linguistic processes of first-person pronoun utterance: " 'Ego' is he who *says* 'ego.' That is where we see the foundation of 'subjectivity,' which is determined by the linguistic status of 'person' " (p. 224). Contrary to the egological tradition, the ego does not precede language but is made possible by it. However, this ego is not constituted in isolation from others, as via a monologue; "I use *I* only when I am speaking to someone who will be a *you* in my address." *I* and *you* are therefore complementary: "neither of the terms can be conceived of independently of the other" (p. 225). Contrary to the egological tradition, a solitary ego is rendered moot, insofar as the practice of discourse necessarily co-involves the one who speaks with the one spoken to. Being complementary, however, does not mean that the two terms are equivalent or symmetrical. Benveniste points to contrast between *I* and *you* – *you* is the non-*I*, the one who I am not and with regard to whom I am transcendent. This difference between *I* and *you*, however, "does not suppress the human reality of dialogue" (p. 201). It is a difference that makes communication *between* persons possible, a difference internal to the dialogic relation. Contrary to the egological tradition, the other person is not construed in a strict symmetrical fashion to oneself as another *I* or ego, which undercuts the possibility of grasping the other from within the context of direct address.

Together with *complementarity* and *asymmetry*, Benveniste views *reversibility* as a key characteristic of the *I-you* relation (p. 199). *I* and *you* are reversible insofar as every *I* uttered by the speaker reverses into a *you* when she is addressed by another and analogously the other person addressed as *you* self-refers by means of the *I*. These *I-you* reversals ongoing within the conversational context are not to be thought as extrinsic to either of the pronouns, as if one could master *I* with no grasp of *you*, and vice versa. As Merleau-Ponty (2000, p. 150) puts it, "The I arises when the child understands that every *you* that is addressed to him is for him an I; that is, that there must be a consciousness of the reciprocity of points of view in order that the word I may be used." The first-person pronoun is grasped in its "fullest linguistic and grammatical meaning" only when the child understands "that even though others call him you he can nonetheless say I" (Merleau-Ponty 2000) Regarded from a developmental perspective, the *I* could never be acquired outside of the context of interpersonal communication, in the solitude of consciousness. Reversibility of *I* and *you* pronouns further exemplifies their complementarity. This *I-you* complementarity is not, however, confined to the early stages of human development when pronouns are acquired; Benveniste emphasizes that a peculiar "polarity of persons," which always posits a *you* as an echo of the *I* (and vice versa), is a permanent trait and fundamental condition of language (p. 225). Discourse is dialogical through and through, and the pronouns provide a privileged locus for witnessing its dialogic dimension. Hence the linguist's assertion that "The

importance of [the] function [of pronouns] will be measured by the nature of the problem they serve to solve, which is none other than that of intersubjective communication" (p. 219). Viewing either of the pronouns in isolation from their inherently communicative role, as is the case in the egological tradition, is therefore to misconstrue the very nature of language, reducing living interpersonal discourse to the sterility of impersonal scientific text.

This dialogical function shared by the first- and second-person pronouns has led some linguists to subsuming them under a unique category of 'interpersonal pronouns' (e.g., Lyons 1979). Benveniste stresses that only *I* and *you* possess the "correlation of personality;" the third-person pronoun, on the other hand, is impersonal or a nonperson (p. 228). If falls outside the scope of direct address and so loses the peculiar discourse-dependent referentiality of *I* and *you*; it refers neither to the one who speaks nor the one spoken to but to the one spoken of, which might be an inert object or a dead person that does not reverse into a *you* or an *I*. Irreversible, static, fixed, the third-person pronoun does not enter the relation with *I*/*you*. Even though it is consistently ascribed to people, it designates them as a nonparticipatory third party, as passive, distant, nonpresent, even though they might be in physical proximity.

Consider some real life examples of the non-identity between the interpersonal *I-you* pronouns and the nonpersonal third-person pronoun, notably the profound transformation that occurs in a shift from direct address to a third-person relation. John Hull, a religious education professor suffering from recent onset blindness reports the unsettling change of attitude among some of his acquaintances. Accompanied by his wife at church, he finds that he is no longer directly addressed but rather spoken about, as when one of the vergers asks his wife, in his presence, "Marilyn, is it John's wish to go forward to the communion rail?" (Hull 2001, p. 101). Even though he appreciates the verger's concern, he is nonetheless upset not to be spoken *to* and notes that "to speak *about* me, in the third person, to someone else, is another matter." Hull feels relegated to the status of objects through the 'does he take sugar?' approach, forcibly removed from the conversational context, even though he is not deaf but blind! His inability to return the gaze is misinterpreted in our predominantly oculocentric society as a linguistic inability to engage in a dialogic exchange, temporarily depriving the blind person of the rights to engage and be engaged as an equal partner in a social setting.[3]

Consider another example, which may for some readers be closer to home. At a cocktail party, you make some pleasant small talk with an acquaintance, and turning to mingle with others, but still in the hearing range of your previous interlocutor, find yourself transformed into a *she* spoken of in a hardly complimentary manner ('She put on weight, did you notice?'). You find yourself objectified, petrified, silenced, your fury is only a testimony of your powerlessness. This moment is reminiscent of the transformation produced by the look of the other in Sartre's famous analysis from *Being and Nothingness*. Engaged in the process of eavesdropping through a keyhole on a conversation unfolding behind closed doors, you find yourself suddenly spotted by another person who happens to pass by in the

hallway. With her gaze she fixes your crouching pathetic figure and reduces you to a humiliating caricature of yourself. For the other, you have become congealed into the manifest façade of an eavesdropper, a role which you now live in the manner of a thing (in Sartre's ontological system, an *in-itself*) rather than as a free project of consciousness (a *for-itself*). Even though the objectification by the foreign gaze temporarily freezes up your freedom, you continue to experience it consciously as humiliation and shame in front of the other. In fact, Sartre argues, the affective states of shame, as well as pride, arise exclusively within the interpersonal world.[4] Note that the Sartrean moment of objectification through the gaze is analogous to the objectification in language through the shift from interpersonal to impersonal pronouns, which effectuates a similar congealment of the subject into a thing. They share the observational or a third-person stance adopted by another person towards you. Even though Sartre contends that this stance typifies interpersonal relations in general, and accounts for their intrinsically conflictual nature, the insights gained from, amongst others, the sociolinguistics of Benveniste suggest rather that they are but one type of dynamics operative within the social world, and one that disrupts rather than preserves the interpersonal relation based on reciprocity, equal partnership, and communicative engagement, within the *I-you* mode.

I concluded the previous section by noting that, insofar as the other is assigned the heading of an *alter ego* within egological phenomenology, the interpersonal relation is construed in third rather than second-person terms. The *alter ego* is grammatically equivalent to a third-person pronoun (as in 'the alter ego thinks…, which can be replaced by s/he thinks…'), and so *functionally* it belongs to the nonpersonal grouping even though *verbally* it claims allegiance to the personal pronouns properly so called. However, insofar as it is cast verbally in the first person only and does not reverse into the second-person form, the *alter ego* is not subject to direct address but must be grasped in third-person terms. As a result of this 'bracketing' of I-you reversal in *transcendentalese*, of which the isolated ego construct is a product, the other person becomes lifted out of the native dialogic context, she becomes *ex-communicated* in the literal sense of the word. No longer a partner in face-to-face communication that I speak *to*, the other person appears as a depersonalized nonparticipatory third party that one may speak *about*, whether casually in gossip or thematically as part of a philosophical project to resolve the problem of other minds. De facto then the alter ego fails to capture the dynamics of direct address in interpersonal relations, even though it may preserve the first-person word in *transcendentalese*.

I conclude that the import of the first-person pronoun into transcendental phenomenology is not a neutral enterprise but that a series of violations of ordinary language principles occurs in the process. Contra Fink, it appears that the natural language is not preserved as is in the transcendental domain, the only difference being the need to read it equivocally. It appears rather that the natural language is disfigured for the purpose of making it say what it would not ordinarily say. The violations include suppressing the *I-you* complementarity, reversibility, and asymmetry, and instituting an isolated *I* defined in terms of autonomy, irreversibility,

and symmetry with other *I*s. These violations motivate the construal of personal identity in terms of a solitary mentalistic subject and of sociality in terms of a detached third-person relation between this subject and analogous others. Egological philosophy commits yet another violation of ordinary language principles when it advocates the use of the ego to refer both to the self and to the other. This usage ignores the 'shifting' character of the pronouns which are not fixedly assigned to self and other in the manner of irreducible transcendental attributes but shift according to the deployment of dialogical roles. This instability of personal pronouns is to be contrasted with the relative stability of proper names which 'stick' to their referents regardless of conversational roles. Negatively put, "There is no concept "I" that incorporates all the *I*'s that are uttered at every moment in the mouths of all speakers, in the sense that there is a concept "tree" to which all the individual uses of *tree* refer" (Benveniste 1971, p. 226). Egological phenomenology does, however, regard the ego as a concept that delineates in the first place the sphere of ownness and applies to other people in a secondary fashion, by analogy to self. From a linguistic standpoint, the ego or the *I* is used as a noun which designates in the first instance one's own self (in the way a proper name does), and in the second instance the other/*alter ego*, rather than as a pronoun which consistently designates the speaker and so refers indifferently to oneself or others depending on the evolving conversational context. As a result of suppressing the shifting nature of the pronoun and instituting a fixed nominal category in its place, egological phenomenology has privileged the self and attributed the kind of ontological primacy to it that is in no way warranted by the ordinary pronoun acquisition and use.

These observations help to counter the objection that the first-person pronoun may be confined to the sphere of ownness even within the context of ordinary language use, insofar as nobody can utter *I* in my place. This professed "inability" of others to substitute me in the process of self-reference is, however, intrinsic to the perspectival character of personal pronouns in general, and as such it should not be regarded as a limitation on the side of either dialogic partner, nor as an exclusive attribute of the first-person pronoun (the second-person pronoun is subject to analogous limitations). Needless to say, the perspectival constraint applies to others as much as to self (I cannot utter *I* in the place of others, yet no transcendental phenomenologist uses this example to postulate the primacy of the other person over the self). The privileging of the self arises only once the dialogic setting of personal pronoun use is undercut and the ego misconceived as a proper name applicable primarily to the self. Consider also that the fact that nobody can say *I* in my place does not imply that I am the only person who can produce – and comprehend – *I*. Nor is it the case that the *I* uttered by self carries a higher semantic load than the I uttered by another person, as if the first-person pronoun pronounced by my conversational partner 'meant less' than the one pronounced by me. And to state the obvious, the production and comprehension of the first-person pronoun is made possible by one's embeddedness in a particular linguistic community – children raised in the German language will typically self-refer by means of *Ich* rather than *I*. This linguistic self-referral does not therefore proceed from inside

out, i.e., from the egological consciousness to the appropriate ego word, but the other way round.

The current discussion of language has direct philosophical relevance (rather than being of a subject matter of linguistic analysis alone) insofar as ordinary language principles capture fundamental aspects of personal identity. To violate these principles is therefore to distort the understanding of what and who we are. Phenomenological philosophy needs to pay heed to how we ordinarily speak in order to render the reality of our lived world without misrepresentation. It will not suffice to go back to the things themselves if we remain forgetful of how things and philosophers are nested within an already existing sociolinguistic fabric which makes the practice of any philosophical inquiry possible.

Due to the series of misconceptions generated by egological philosophy regarding the person and interpersonal relations, the notion of the transcendental ego needs to be questioned and ultimately abandoned. In fact, Sartre reached a similar conclusion as early as 1936 (Sartre, 1972). In his view, phenomenology must declare the ego to be transcendent to the field of consciousness uncovered by the epoche if it is to definitively remove the threat of solipsism haunting transcendental phenomenology. The transcendental ego is the stumbling block on the path of social phenomenology, making it conceptually and not only empirically impossible to thematize relations with other conscious minds. Sartre's own alternative proposal was not to reject the notion of the ego *tout court* but to regard it as a mundane-like thing, transcendent to the domain of pure consciousness and available in equal measure to the self and to others. To be sure, it is highly debatable whether this alternative succeeds or fails to advance the phenomenological theory of social relations, especially since Sartre preserved the notion of consciousness as inner and private mental arena, accessible exclusively in the introspective mode. The author himself admitted failure in this regard (Sartre 1956). His starting intuition that phenomenology must renounce the egological base if it is to do justice to the reality of sociality retains, however, its full force. This intuition is shared by the philosophy of dialogue, and, as argued below, the latter may be in a better position to provide adequate theoretical tools for social phenomenology than the Sartre of *The Transcendence of the Ego* has been.

The analysis carried out up to this point raises the question of the exact importance of ordinary language for the study of personhood and interpersonal relations. If we follow Benveniste to the letter, we would need to view persons as exclusively linguistic entities, whose existence is coextensive with each instance of discourse and thus, of a "momentary" nature (1966, p. 226). This exclusively linguistic construal of persons encounters a range of problems of its own: Do persons blink in and out of existence? Are infants (from Latin *infans*, literally speechless) not persons? What about the deaf and mute, the aphasiacs? Verbalized discourse cannot fully capture personhood, and the demise of structuralism is a historical proof of the inadequacy of theoretical explanations confined to linguistic models. However, the view at the opposite end of the spectrum, embraced in part at least by transcendental phenomenologists, which regards ordinary language as contingent and extraneous to persons and interpersonal relations, is equally problematic.

To name just one difficulty, verbal competence is turned into an unprecedented accomplishment that arises *ex nihilo* in the human world. Consequently, no story of gradual emergence of linguistic from prelinguistic skills could be told and no continuity between preverbal and verbal communicative strategies could be traced, whether in evolutionary or developmental contexts. Contemporary research in developmental psychology challenges this disjunction, and since the focus of this paper is on personal pronouns, I will limit the discussion of continuity between preverbal and verbalized communicative strategies to pronouns. This is not, and I stress this point emphatically, to limit the study of relations between persons to the study of personal pronouns acquisition. It is, however, to insist on the continuity between preverbal and verbal communicative skills, and on the fact that the latter must be included in a complete account of how persons communicate.

5.3. DEVELOPMENTAL PERSPECTIVES

Katharine Loveland (1984) researched this interspace spanning prelinguistic and linguistic skills operative within human communication. Specifically, she pioneered a series of studies which demonstrate that the acquisition of first and second person pronouns and their possessive forms *my, mine, yours,* taking place around the age of three is correlated with the understanding of spatial perspectives. As she notes, *I* and *you* are always used relative to the speaker's own point of view and their referent shifts depending on the context of utterance. Acquiring an understanding of this shifty nature of pronouns may be a long and laborious process, slowed down by the so-called reversal errors where the child fails to appropriately reverse in the production of *I* and *you* (as in the exchange 'Do you want a cookie, Annie? Yes, you want a cookie"). The acquisition of the first and second person pronoun deictics depends therefore on a developed nonegocentric stance (in contradistinction to the noted egocentrism of the transcendental ego). It is also noteworthy that autistic children, whose difficulties in navigating the social space have been well documented, encounter severe difficulties in managing first and second pronouns correctly due to problems with grasping the reciprocal relations involved in their use (Loveland 1993). Finally, there is some evidence that blind children are delayed in the *I/you* acquisition, possibly due to the delay in appreciating differing personal perspectives.

These difficulties lend credence to the hypothesis that a perceptual/spatial appreciation of points of view is a necessary prerequisite of pronominal competence involving *I/you*. Loveland tested this hypothesis in a series of cross-sectional and longitudinal experiments. In the former series, the experimenters engaged a group of children in a variety of spatial tasks, such as conspicuously hiding a toy and requesting that the child retrieve it, conspicuously shifting the gaze in one direction to elicit gaze following in the child, requesting the child to show a picture to Mommy so that she can see it. These spatial tasks centered on understanding differing points of view were followed by language tasks, which tested pronominal competence in the *I/you* comprehension and production. For example, the adult asked 'What do

I/you have?' when both have a toy, but not the same one. In other examples, the adult asked 'Who has the toy?' when either adult or child has it, and 'Whose toy is it?' when it is either adult's or child's. The results confirmed the hypothesis that only children who demonstrated understanding of differing points of view in the spatial tasks were able to produce and comprehend *I/you* pronouns correctly.

The longitudinal experiments studied how spatial point of view and reciprocal deictics interrelate over the course of development. In these experiments, the author also found that progress in acquiring the *I/you* pronouns was correlated with increasing knowledge about points of view. She concluded that there is a clear link between the child's growing understanding of the visual/spatial relations among persons and the child's emerging ability to refer to persons pronominally. Loveland's results have since been confirmed and extended to cover the comprehension and production of the third person pronoun by Ricard et al. (1999). The authors confirmed that children's performance on perspective-taking tasks was requisite to full pronoun acquisition. Interestingly, they demonstrated that competence at coordinating two visual perspectives preceded the full mastery of the first and second person pronoun, and competence at coordinating three visual perspectives preceded the full mastery of the third person pronoun in a nonaddressee context. As the researchers were careful to note, early perspective-taking skills are not sufficient to explain all aspects of personal pronoun acquisition. Phonological and syntactical abilities must also be taken into account in the full story of pronominal competence. However, perspective-taking skills in spatial context do contribute to the perspective-taking skills in the linguistic context of pronominal reference. Spatial and linguistic perspectives are correlated in the cognitive development of the child: Both involve an understanding of persons as perspectively oriented, whether this orientation attaches to spatial or linguistic anchoring. We gather from the preceding that perspectivity is key to the notion of personhood, and that it serves as a bridge connecting the prelinguistic perceptual and linguistic grasp of persons, whether they are defined visually/spatially as seers or onlookers (themselves visible) or linguistically as speakers/addressees. Visually defined, persons negotiate a common public space by means of the gaze which can be directed at one another in eye contact or aimed at a third term. An abundant body of research has investigated these visual/spatial skills in the light of their social significance of establishing mutual contact and joint attention mechanisms.[5] They belong to the wide category of the so-called *protoconversation*, which centers on the conversational-like patterns of give and take manifest in infant and caregiver dyads, including not only the visual but also facial and vocal contingently structured exchanges. These protoconversational abilities, which serve as social glue for the dyad before language comes on board, retain their communicative potential throughout human life. Persons converse, so to say, with their voice, gaze, face, and body as much as with the articulated word, and dialogical exchange needs to be construed broadly to account for all these correlated forms of second-person interaction, as it has consistently been in ecological psychology.[6]

Whereas visually defined persons negotiate a common public space, linguistically defined persons negotiate a common dialogical space by deploying the roles of speaking and listening one to another, including speaking about a third. Persons are therefore not produced *ex nihilo* by personal pronouns and pronominal competence is not the starting point of interpersonal communication. Personal pronouns do, however, draw on and deepen our grasp of both the plurality and the interrelation of individual viewpoints, and so need to be included in the philosophically satisfactory account of sociality. Misconstruing pronouns does lead to a partial misconstrual of what persons are.

5.4. ADDITIONAL CONTRIBUTIONS: THE DIALOGIC TRADITION, ECOLOGICAL PSYCHOLOGY

In conclusion, I propose to briefly outline other existing intellectual resources, which establish the primacy of second-person relations. In doing so, I am following the lead of Theunissen's classic study *The Other* (1984). Theunissen argued that Husserl's transcendental project is oriented by the preconceived agenda in the direction of world constitution; as a result, his entire analysis of sociality is subsumed to this predetermined aim, and the experiential givens of social life are misconstrued. If that is the case, we should search for a philosophical alternative that is free of such preconceived agendas in its interrogation of sociality. Theunissen advocates resorting to the tradition of dialogical philosophy, exemplified notably by Buber, Ebner, Marcel, Rosenstock-Huessy, and Rosenzweig, which preserves the primacy of the relation between *I and you*.[7] In the following, I will briefly outline the primacy of this relation by focusing on relevant sections from Buber's *I and Thou*.

Like Loveland and Benveniste, Buber does not regard the *I* as an independent entity. "There is no I as such," but rather two kinds of basic couplets in which I exists: the I-you and I-it (p. 54). While the former denotes a relation of reciprocity, the latter is founded on a unilateral dependence of an object on the subject. Both stances may involve other human beings as well as inhabitants of the natural world (animals, plants, even rocks). The designation as *you* or as *he/she/it* does not therefore reflect an intrinsic quality of the referent; instead, it manifests the way of relating to it, either by means of an unmediated openness with one's whole being that occurs in direct address or by means of detached observational and instrumental approach. In the former case, one reaches out to "nothing," i.e., one does not experience a discrete qualitatively defined entity but rather catches a glimpse of the eternal; in the latter, one attains a being as an aggregate of qualities which can be defined as such and such, and used as a means to an end. This duplicity of approach to others is reflected within the twofold nature of *I* as it engages in a relation of reciprocity or functions as the subject of objectification (p. 53). Only the *I* of the *I-you* relation is a person, while the *I* of the *I-it* relation lacks the dimension of personhood (it is an *Eigenwesen*, translated as an ego, p. 111).

The distinction between *I-you* and *I-it* relations reflects therefore Benveniste's distinction between interpersonal and nonpersonal pronouns, where the correlation

of personality applies exclusively to *I-you*. Both authors view personhood as dialogically constituted within the context of direct address rather than as an intrinsic quality belonging to certain beings. That is why annexing humans with mind and interiority does not lift them out of the domain of it; in Buber's words, "Inner things like external things, things among things!" (p. 56). It is only as an *I* in a relation to a *you* that a being acquires personhood and self-consciousness ("the consciousness of the constant partner, the I-consciousness", p. 80), and reaches beyond the domain of nature to the domain of spirit.[8] In its relation to others, a third-person approach confines humans to the status of nonpersons.

Like Benveniste, Buber stresses the complementary of *I* and *you*: *I* emerges only in relation to *you*, not prior to it (p. 80). Furthermore, for Buber, as for Benveniste, *I* and *you* are strictly asymmetrical and it would be misguided to thematize *you* as though it were a projected image of *I*. Such a relation would disintegrate the interpersonal relation into an I-it stance, despite the warmest feelings one could have for 'another I' (Friedman 2002, p. 70). Unlike Benveniste, however, Buber does not explicitly discuss I-you reversibility, and the reader of *I and Thou* gets the sense that *I* is viewed as synonymous with self, while *you* with another person. Even though I cannot expand upon this point in the present essay, such a fixing of *I* and *you* may ultimately run contrary to the efforts to decentralize the self.

Benveniste and Buber share the basic assumption about the primacy of interpersonal relations in the *I-you* mode. While Benveniste regards it exclusively as a linguistic relation dependent on the instance of discourse, Buber enlarges it to nonlinguistic forms of reciprocal engagement which may involve shared gaze and touch, and so are operative within infancy as well as adulthood (p. 79). Contra Benveniste, language acquisition is therefore not a precondition for entering the I-you relation; rather an "*a priori* of relation; the *innate You*" informs the infant's earliest gestures and vocalizations as much as it plays out in the verbalized exchanges of later life (p. 78). Furthermore, for Buber the utterance of the lexical *I* and *you* alone does not guarantee that a genuine relation of dialogue has taken place; one may use the words and yet be caught in an *I-it* type of relation (p. 118). I conclude that Buber liberates the I-you relation from an exclusively linguistic construal and provides a richer conception of interpersonal relations that correlates, to borrow Trevarthen's (1993) term, *proto*conversational strategies with conversations properly so-called. Buber establishes the I-you relation as an irreducible ontological dynamic which is not a subject matter of linguistic analysis alone but belongs to the domain of philosophy. As such, he provides a veritable alternative to the philosophy of sociality from the egological tradition. He enables us to move beyond the constraints of a *monologically* defined phenomenology of consciousness to the phenomenology of the *dialogical relation*, or to coin an appropriate term, to dialogical phenomenology.

I need to clarify that by advocating dialogical phenomenology, I do not call for a replacement of phenomenology by the philosophy of dialogue. This view, which could be termed the *revolutionary* stance, may have been embraced by Theunissen.[9] Its disadvantage is that it cuts itself off from the rich resources provided

by phenomenological philosophy, notably regarding embodiment. Another possible approach, which – to stick to political metaphors – can be termed *coalitional*, would be to combine some of the dialogical contributions with the phenomenological project without revising its egological base. This approach is exemplified by the phenomenological sociology of Alfred Schutz (1967), notably his analysis of the Thou-orientation carried out in the framework of ego-centered phenomenology.[10] Merleau-Ponty (1973, 1994) carried out a similar, albeit far less extensive, project of thematizing sociality from the egological perspective while incorporating dialogical insights. This project descriptively examines interpersonal dialogical relations while maintaining the theoretical framework within which the other is construed as a replica of the self – or another I. Both approaches are pervaded by an inescapable tension between the two philosophical traditions that do not share their starting points and so do not marry happily – unless the partners be transformed in the process, just like the partners may be transformed in the process of dialogical exchange. In that case, the advocated method of study would reflect the phenomenon under investigation and both the method and the social phenomenon would involve a dialogical exchange. That is why I would like to advocate yet another approach, which will be termed a *revisionist* stance.

This approach does not only combine key insights of both phenomenological and dialogical disciplines, but also revises some of their problematic assumptions. This revision would be accomplished by engaging the two disciplines in a relation of mutual critique and enlightenment. On the one hand, this stance advocates relinquishing the egological starting point for the sake of the primacy of the I-you relation. On the other hand, it advocates enriching the notion of the dialogical relation notably with the phenomenological contributions concerning the inescapable embodiment of dialogical partners, as found especially in the work of Merleau-Ponty and Sartre. I believe that this approach may provide impetus to phenomenological research and exploit the potential of phenomenological description in the area of direct social interaction. Furthermore, contributions from ecological psychology, such as the aforementioned protoconversational dynamic operative in infancy, as well as the currently actively researched question of language acquisition in its dependence on perceptual modes of social interaction (esp. Tomasello 2005) will play a vital role in further fleshing out the dialogic relation. I hope that the analysis carried out in this chapter testifies to the dire need of pursuing this intellectual project, as well as provides some cues for how to enact interdisciplinary dialogical phenomenology. I believe that the latter may provide a better alternative to the so-called folk psychology-based models of social cognition than the egologically informed phenomenological accounts ever could. Last but not least, I hope that the analysis offered here, especially the material from linguistics, helps to demonstrate the wrong headedness of the folk psychological approach to social cognition. By focusing on the other construed as a nonparticipatory third party, the defendants of folk psychology are choosing to focus on instances of social observation and detachment and are necessarily by-passing the fundamental structure of intersubjectivity, i.e., social interaction in the second-person mode.

They overlook not only the developmental data on the primacy of face-to-face interactions, but also the linguistic and phenomenological data on the primacy of I-you connectedness over an impersonal stance towards *she, he,* or *it*. I hope to have shown that phenomenological description coupled with linguistic analysis provides a powerful tool with which to debunk the key presuppositions of folk psychology regarding social life. However, as I argued, egological phenomenology might not be the best place to look.

ACKNOWLEDGEMENTS

My thanks to Mark Johnson, Dan Hutto, Matthew Ratcliffe, as well as to Sarah Adams for helpful comments on early drafts of this chapter.

NOTES

[1] Benveniste discusses specifically how 'being' has become part of the Western philosophical idiom.

[2] In Strawson's own words, "We are tricking ourselves by simultaneously withdrawing the [first person] pronoun from the ordinary game and yet preserving the illusion that we're still using it to play the ordinary game."

[3] I discuss further the implications of eye contact for interpersonal interaction in Stawarska (2006).

[4] I expand on Sartre's analysis of the effect of another person's gaze in Stawarska (2004).

[5] See e.g., Butterworth and Cochran (1980) and Butterworth and Jarret (1991).

[6] On protoconversation, see especially Trevarthen (1993).

[7] Theunissen's *The Other* typically uses Thou in place of you. For the sake of consistency with preceding sections, I use *you* throughout. This usage is consistent with Buber (1970, p. 14) who noted that Thou is a primarily theological term, while *you* can capture the spontaneity and unpretentiousness of human relations.

[8] Again, the spirit for Buber is nothing internal. "Spirit is not in the I but between I and you" (p. 89).

[9] See especially the "Postscript: The Transcendental Project of Social Ontology and the Philosophy of Dialogue" in Theunissen (1986).

[10] Discussed by Theunissen (1986) in Appendix 1. "Transcendental Philosophy and the Illusion of Dialogue: Alfred Schutz's Illusion of Dialogue."

REFERENCES

Benveniste E (1971) Problems in general linguistics. University of Miami Press, Coral Gables.

Bogdan R (1997) Interpreting minds: the evolution of a practice. MIT Press, Cambridge, MA.

Buber M (1970) I and thou. Touchstone, New York.

Butterworth G, Cochran E (1980) Towards a mechanism of joint visual attention in human infancy. Int J Behav Dev 3:253–272.

Butterworth G, Jarret NLmM (1991) What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. Br J Dev Psychol 9:55–72.

Fink E (1995) Sixth Cartesian meditation. Indiana University Press, Bloomington.

Friedman M (2002) Martin Buber: the life and dialogue, 4th edn. Routledge, London.

Gallagher S (2001) The practice of mind: theory, simulation, or primary interaction? J Conscious Stud 8:83–108.

Gurwitsch A (1979) Human encounters in the social world. Duquesne University Press, Pittsburgh.

Hull J (2001) On sight and insight. One World, Oxford.

Husserl E (1970) The crisis of European sciences and transcendental phenomenology. Transl. by Carr D. Northwestern University Press, Evanston.
Husserl E (1991) The Cartesian meditations. an introduction to phenomenology. Transl. by Cairns D. Kluwer Academic Publishers, Dordrecht, Netherlands.
Husserl E (2002) Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book. Kluwer Academic Publishers, Dordrecht, Netherlands.
Hutto D (2004) The limits of spectatorial folk psychology. Mind Lang 19(5):548–573.
Jakobson R (1973) Essai de Linguistique Generale. Les Editions de Minuit, Paris.
Loveland K (1984) Learning about points of view: spatial perspective and the acquisition of 'I/you'. J Child Lang 2:535–556.
Loveland K (1993) Autism, affordances, and the self. In: Neisser U (ed) The perceived self. Ecological and interpersonal sources of self-knowledge. Cambridge University Press, Cambridge.
Lyons J (1979) Deixis and Anaphora. In: Myers T (ed) The development of communication and discourse. Edinburgh University Press, Edinburgh.
Merleau-Ponty M (1973) Dialogue and perception of the other. In: The prose of the world. Northwestern University Press, Evanston.
Merleau-Ponty M (1994) The phenomenology of perception. Routledge, London.
Merleau-Ponty M (2000) The child's relations with others. In: The primacy of perception. Northwestern University Press, Evanston.
Ricard M, Girouard P, Gouin Decarie T (1999) Personal pronouns and perspective taking in toddlers. J Child Lang 26:681–697.
Sartre JP (1956) Being and nothingness. An essay on phenomenological ontology. Philosophical Library, New York.
Sartre JP (1972) The transcendence of the ego. Octagon Books, New York.
Schutz A (1967) The phenomenology of the social world. Northwestern University Press, Evanston.
Stawarska B (2004) The body, the mirror and the other in Merleau-Ponty and Sartre. In: Gallagher S, Watson S (eds). Ipseity and alterity: interdisciplinary approaches to intersubjectivity. Presses Universitaires de Rouen, Rouen, pp 175–186.
Stawarska B (2006). Mutual gaze and social cognition. J Phenomenol Cogn Sci 5(2):17:30.
Strawson PF (1958) Persons. In: Feigl H, Scriven M, Maxwell G (eds). Minnesota studies in the philosophy of science, vol. II. University of Minnesota Press, Minneapolis, pp 330–353.
Strawson PF (1966) Self, mind, and body. Common Factor 4:5–13.
Theunissen M (1986) The other. Studies in the social ontology of Husserl, Heidegger, Sartre, and Buber. MIT Press, Cambridge, MA.
Tomasello M (2005) Constructing a language: a usage based theory of language acquisition. Harvard University Press, Cambridge, MA.
Trevarthen C (1993) The self born in intersubjectivity: the psychology of infant communicating. In: Neisser U (ed). The perceived self. Ecological and interpersonal sources of self-knowledge. Cambridge University Press, Cambridge.

PART II

REASONS, NORMS, NARRATIVES AND INSTITUTIONS

PETER GOLDIE

6. THERE ARE REASONS AND REASONS

6.1. INTRODUCTION

Why do people do the things that they do? This is a very general question, and I want here to treat it as such, without unduly narrowing it down. I do, however, want to restrict the question to those things done that are intentional actions, or things done for a reason in the particular sense that Elisabeth Anscombe was searching for in her book *Intention*.[1] People sometimes do things other than what we would consider to be intentional actions. When Talleyrand (the great French diplomat who served in turn the Ancient Régime, the Revolutionary government, Napoleon I, the restored monarchy with Louis XVIII, and Louis-Philippe after the 1830 revolution) finally died in 1838, Metternich famously remarked 'I wonder what he meant by doing that'. But people do not die intentionally, for reasons in the sense required for meaningful intentional action (suicide is something quite different), and that is what makes the remark into a nice compliment to Talleyrand, who was said never to do anything without good reason.

Now, following on from Anscombe, and since the work of Donald Davidson, I will characterise intentional action as action done for a primary reason, consisting of a belief and an attitude towards this kind of action (I will from now on call it a desire, in the knowledge that this term is desperately vague; but nothing hangs on it here), and this belief and desire will give the answer why, in the sense we want, people do the things that they do. As Davidson puts it, the primary reason rationalises the action. And it will also causally explain it.[2]

That intentional action can be explained in this way, by reference to mental states of the individual – beliefs and desires – is now pretty much a philosophical commonplace, and the idea now has claim for a kind of monopoly. In fact, it has claim for two kinds of monopoly. First, it claims a monopoly in the sense that all intentional actions are supposed to be explainable by appeal to the individual's beliefs and desires – by what I will from now on call a belief-desire explanation. This first monopolistic claim has recently been challenged, but I have no quarrel with it so far as it goes;[3] in fact, I think it is correct.

It is the second claim for monopoly that is my target. This is the claim that belief-desire explanation fully and satisfactorily explains intentional action, or, alternatively, that any other kind of explanation will ultimately resolve into a belief-desire explanation. When it comes to action explanation (and prediction too for that matter, although that will not be my concern here), the belief-desire story is supposed to be the only game in town. For example, Jerry Fodor uses the terms 'commonsense psychology' and 'commonsense belief-desire psychology' pretty much interchangeably, and Donald Davidson says that

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 103–114.

constructing a primary reason (a belief-desire pair) is not only necessary but also sufficient to rationalise an action. More recently, Greg Currie and Ian Ravenscroft place all their emphasis on the idea that our 'everyday understanding of minds' requires a grasp of 'the beliefs and desires of someone …whose behaviour we want to predict or understand', and Shaun Nichols and Stephen Stich say that 'the central concepts implicated in mindreading' are 'belief, desire, intention'.[4]

In challenging this second claim for monopoly, I will try to show that our everyday explanations of intentional action – as part of our so-called folk psychology – are characteristically quite different from this. What will emerge, I hope, is that these everyday kinds of explanations of action, which I will call *thicker explanations*, are much more revealing, much more far-reaching, and much more varied than belief-desire explanations. Because I am not challenging the first monopolistic claim, I am not denying the availability of belief-desire explanations. What I am denying is their explanatory adequacy; belief-desire explanations are seldom sufficient *as explanations*. Moreover, in their detail belief-desire explanations are not necessary either, except in special circumstances of the kind I will be discussing.

If this is true, then there are significant implications for the philosophical debate concerning the way we go about our everyday explanations of other people's actions. If the belief-desire story was the only game in town, then any explanation of action would have to appeal to these reasons – beliefs and desires as occurrent mental states of the individual – and appeal to these reasons would be sufficient to explain an action. And then the question becomes pressing as to how we 'gain access' to, or come to know, another person's beliefs and desires. Do we gain a grasp of these reasons, these occurrent mental states of the other person, by theorising about them, as what are sometimes called 'unobservables'; or do we simulate or imaginatively project ourselves into the shoes of the other person in order to generate imaginative counterparts of the other's mental states in our own minds, and then assume that the other is thinking as we are? It is now fashionable to call this the 'mindreading' debate, which is the term supposed to bring out the apparent mysteriousness of how we can 'gain access' to what is going on in another's mind.[5] But if my claim is true, then this whole debate would seem to have the wrong focus: 'Mindreading', in the sense of 'reading' the goings-on in another's mind, is only a small part of what is necessary for our everyday framework of action explanation, and this is because the detailed goings-on in the other person's minds as causes of the action are not relevant to those thicker explanations that I will be discussing. What the current debate seems to have lost track of is something that ought not to be at all controversial, namely the sheer complexity of the aetiology of intentional action, and the consequent diversity of possible causal factors that can be appealed to in any given explanation. Beliefs and desires, as occurrent mental states, are only one causal factor, and appealing to them to explain an action is (except in special circumstances) redundant.

6.2. BELIEF-DESIRE EXPLANATIONS AND THIN RATIONALITY

In our everyday discourse about action explanation, when do we refer to beliefs and desires? The kinds of example that are discussed in our lectures in the philosophy of mind are usually very simple. Someone opens the door of the fridge because he wants a beer and believes that opening the door is the best means of getting a beer, believing, as he does, that there is a beer in the fridge. Someone takes an umbrella because she believes that it is going to rain later on, because she wants not to get wet and because she believes that taking an umbrella is the best means of avoiding getting wet.

Then the lecturer goes on to point out that our explanations sometimes only refer to a desire ('he wanted a beer') and sometimes only to a belief ('she thought that it was going to rain'), but this is merely for Gricean pragmatic reasons, because reference to the other mental state is redundant, given the explanatory needs of the situation; it does not imply that a full belief-desire explanation is not available. It is also pointed out that this kind of belief-desire explanation can be expanded to allow for much more complex examples of intentional action.

I think this is all fine; in respect of all intentional actions, such belief-desire explanations are indeed available. This is the first monopolistic claim, and I am not disputing it. Indeed, it seems to me that in our thinking about action, our own and other people's, we just *take it for granted* – it is, I dare say, a priori – that belief-desire explanations are always going to be available.

Although we take it for granted that belief-desire explanations of intentional action are always going to be available, in practice we really only need to turn to their detail under special circumstances. Sometimes we turn to them when there is some kind of mistake involved, and in particular when one or more of the beliefs involved is false: There is no beer in the fridge. Sometimes we turn to them for forensic reasons; we might ask if she actually believed that this umbrella that she took was hers, or whether she just took the first umbrella that she saw. On such occasions, it is the specific details of what is in the person's mind that matters.

Where these special circumstances do not obtain, the reason why we do not turn to belief-desire explanations is that they are so thin as to be redundant or of little explanatory use. They explain so little partly because they rely, in turn, on such a thin notion of rationality. To rationalise an action in this thin sense is just to show how it could make sense for someone to do such a thing. And it can often make sense to do one of a diverse range of possible things. Someone is the non-paying guest in a restaurant, and the waiter brings him the wrong flavour of ice cream, explaining that they have run out of the chocolate flavour, which he ordered, and that this is why the waiter has brought him the strawberry flavour instead. It would make sense for one to tell the waiter to take it back; it would make sense for one to eat what one is given; it would make sense for one to leave it uneaten; it would make sense for one to throw the ice cream on the floor and walk out of the restaurant; it would make sense for one to offer the ice cream to one's host; it would make sense for one to pour the ice cream onto one's host's lap, etc. So we must not forget that a belief-desire explanation does not explain why someone did

one thing rather than another which was also open to him and which also would have made sense. The point is at its most stark if you are asked to *predict* what a rational person (that is all we know about him or her) will do if he or she is brought the wrong flavour of ice cream by a waiter in a restaurant. 'Well', you will sensibly reply, 'it depends'.[6]

This thinness gives a supple kind of strength to belief-desire explanation, for it leaves the explanation open to be thickened out or supplemented in all kinds of ways, with reasons other than beliefs and desires.[7] Assume that our diner left the ice cream on his plate. To a rough approximation, the belief-desire explanation is that he did this because he desired not to eat the ice cream, believing it to be strawberry-flavoured, and he desired chocolate flavour and not strawberry. But this is singularly uninformative. It is the thicker explanation which is more interesting to us and which will give us the kind of interesting and revealing explanation that we are seeking when we ask why he left his ice cream. For example, does his action reveal a certain inconsiderateness towards his host?

It is to the various kinds of thickening out that I now turn. I will put them into four broad categories – more for the sake of discussion than out of some taxonomising zeal. The first kind of thicker explanation is concerned with motive and desirability characterisation.

6.3. THICKER EXPLANATIONS

6.3.1. Motive and desirability characterisation

I will start with Elisabeth Anscombe's famous example of 'the man who (intentionally) moves his arm, operates the pump, replenishes the water supply, poisons the inhabitants'. Anscombe says that the man is performing just one action which has four descriptions, 'each dependent on wider circumstances, and each related to the next as description of means to end'…'the last term we give in such a series [A, B, C, D] gives the intention *with* which the act in each of its other descriptions was done, and this intention so to speak swallows up all the preceding intentions *with* which earlier members of the series were done. The mark of this "swallowing up" is that it is not wrong to give D as the answer to the question "Why?" about A'.[8]

The thinnest of belief-desire explanations in the example is that the man wants to move his arm and believes that he can do this by doing just what he is doing. It is less thin if the explanation appeals to the ultimate intention, the last in the series, which is to poison the inhabitants; so this belief-desire explanation would then be that the man wants to poison the inhabitants and believes that he can do this by doing what he is doing, namely operating the pump by moving his arm in this way. But our interest in such a case extends beyond this, to find an explanation of *why* the man wants to poison the inhabitants. We are looking beyond his desire, for some *desirability characterisation* as Anscombe calls it, which will capture what is desirable for this man about poisoning the inhabitants. Let us now assume that 'The bastards deserve to die in agony after what they did to us' captures this desirability characterisation.

This seems now to be an act of revenge. Revenge was the *motive*. As Anscombe says, 'popularly, "motive for an action" has a rather wider and more diverse application than "intention with which the action was done"'; '[t]o give a motive …is to say something like "See the action in this light"'.[9] I would like to put it like this: The notion of motive is both summarising and evaluative of what is in the mind of the person doing the action, but it does not imply that the motive, as such, was in the person's mind as he did the action. So when we say that this man's motive was revenge, we do not imply that revenge as such was in his mind as he was operating the pump. The man might be so caught up in the desperate cycle of revenge that he might not even dwell on the desirability characterisation of his action; he simply carries on, tit for tat, atrocity for atrocity. A man can be 'set on' revenge but not be thinking about revenge as such. The point in general, though, applies equally to motives such as curiosity, spite, friendship (examples of Anscombe's), vanity, pride, self-improvement, ambition, greed, envy, modesty, inconsiderateness.

Take the first of these: an action done out of curiosity. You have invited your neighbours round for drinks. We see one of them wander out of the drawing room, heading up the stairs towards the bedroom. I ask you why she is going up the stairs, and you reply 'Curiosity'. If it is true that she is doing this out of curiosity, then the appropriate desirability characterisation might be 'It would be interesting to see what their bedroom looks like'. Perhaps if pressed as to why she thought she would find that interesting, she might say, 'Well, come to think of it, I suppose I was just curious,' but that is not to give a further reason – a further thought, latent or occurrent – to causally explain her action; rather it reveals that she later accepts that it is appropriate for us to see her action in this light. In giving this thicker kind of explanation, appealing to motive, we go beyond the belief-desire explanation.

Now consider the last of the motives on my list: inconsiderateness. A man opens the fridge door, looks in, sees just one remaining slice of lemon pie and eats it. Let us say that his only relevant desire was for food (a so-called unmotivated desire[10]), and, if pressed, the best he might be able to say to explain his action is 'Well, I was just hungry, and there was the lemon pie.' But our thicker, evaluative, explanation claims that he was being inconsiderate, and it goes further than the belief-desire explanation: It accepts the truth of the belief-desire explanation, and adds the idea that something was *lacking* in this man's motivations: he failed to take account of others' interests (we might also have said he was being selfish and thoughtless). This explanation of his action – 'he was being inconsiderate' – is a kind of evaluatively loaded causal explanation: Just as we say that the building collapsed because the builders did not bolt the beam to the wall (as they should have done), we say that the man took the last slice of pie because he failed to consider others' interests (as he should have done). In saying that he was being inconsiderate, there is no implication that his thinking involved first considering others' interests, and then on reflection he found them to be of less importance; on the contrary, the remark suggests that the interests of others never so much as entered his mind. So this example shows clearly that one cannot always equate motive and intention. And

how much more interesting and revealing is this thicker explanation than the one that just cites the man's actual beliefs and desires – what actually went on in his mind at the time!

In thickening out the explanation by appeal to motive and desirability characterisation, we also leave beliefs and desires behind in this sense: Having determined what the motive and desirability characterisation were, we are not interested in the *detailed* goings-on in the mind of the person doing the action, the *specific* beliefs and desires, except under special circumstances of the kind I mentioned earlier. Moreover, the order of discovery does not have to be, and typically is not, first discovering the specific beliefs and desires and then inferring the motive and desirability characterisation. Sometimes the motive can be manifest in a person's intentional action – consider greed and vanity as examples. On other occasions, the process can be one of testing various possible motives and desirability characterisations against the specifics of particular actions. Detectives work like this. What was his motive for suffocating his terminally ill wife – love, financial gain, revenge or loathing? But if it had been financial gain, why did he show such complete lack of interest in her financial affairs? Thus, we can test each possible motive against the detail, but finally it is the motive that concerns us more, not just because it is thicker than the belief-desire explanation, but also because it is evaluative.

Motive and desirability characterisation, then, are the first kind of thicker explanation of action, beyond what is going on in the person's mind when he acts. The second is character and personality.

6.3.2. Character and personality

We often appeal to someone's character or personality to explain their action. Character traits, as well as many personality traits, are, roughly, dispositions reliably to have (or to lack) certain kinds of motives in certain kinds of situation, and thus reliably to act in certain kinds of ways. Although some personality traits, for example habits and action tendencies such as being fidgety are simply dispositions to behave in certain ways and in certain kinds of situation, and are not concerned with motive, it is a profoundly serious mistake to assume that all traits are merely dispositions to behave.[11] Most involve motive. To appeal to one of these traits to explain an action implies a certain kind of motive, but it goes further: It also implies a certain degree of reliability. 'Why did he eat that last slice of lemon pie?' The reply, 'Because he's inconsiderate' goes further than 'Because he was being inconsiderate'; the former implies both lack of consideration at the time and a reliable lack of consideration in this kind of situation; the latter merely implies a lack of consideration at the time.

Appeal to character and personality traits in this way certainly is a kind of action explanation. But it is not an explanation that cites the trait itself as a cause; rather, it points towards some sort of motive as a cause. The explanation that she helped him pick up his papers after they had fallen out of his briefcase because she is a kind and helpful person implies, first, that she has a disposition reliably to have kind and helpful motives when appropriate, and secondly, that this disposition was

operative on this occasion: she did have these motives. Of course in explaining an action in this way one does not pretend to explain why she has a kind and helpful disposition. It is, you might say, a dormitive virtue explanation, and these explanations have their place in our explanatory framework.[12]

I should mention here a claim that is sometimes made by philosophers and social psychologists: That, whilst it is accepted that our everyday psychological talk is replete with talk of character, there really are no such things as reliable character traits, understood as robust traits that are both stable and consistent across a broad and diverse range of situations.[13] I have discussed these claims elsewhere, so my comments here will be brief.[14] The first comment concerns the stability of traits. Our everyday psychological practice crucially depends on appeals to stable motives – in other words to mental dispositions, and not just to occurrent mental states. He leaves work early to go to his son's school play, in spite of the important deal that had to be done that day. When we explain this by saying that he left early because he is a loving parent, the explanation does more than appeal to his occurrent loving thought about his son. It appeals to his loving disposition towards his son, of which his occurrent loving thought is an expression. Now, it is not generally denied here that there can be such dispositions; so let us assume that he has such a disposition. The question then turns to consistency, which is where the pressure on traits generally rises. It is true that we do expect consistency in respect of a trait such as love of one's son, which ought to find expression in all kinds of situation, and not just when it is time to go to the school play. And it is also true that we are often disappointed: We can disappoint ourselves and others by acting out of character; and social psychologists have shown great ingenuity in proving it. But what this shows (so I argue elsewhere) is not that we should drop all talk of character, and of virtue, because character and virtue fail to manifest themselves with the required consistency across a diverse range of situations. Rather, it shows that our talk of character is *idealistic*. If we say that someone is loving towards his children, then we imply that he, as such a person, *ought* to have loving thoughts towards his child in all sorts of different situation, and that he *ought* to act accordingly. And of course it is possible to fail to have the thoughts and to do the things that one ought to do as a loving person whilst still being a loving person. I will turn to an example shortly.

Explanations of action in terms of character and personality, then, also go beyond belief-desire explanations. They indicate some kind of stability and consistency in motive and action. They are evaluative and normative. And – like motive in this respect too – they are less specific than belief-desire explanations. They just point towards a general sort of motive (or lack of motive) – kindness, helpfulness, inconsiderateness, and so on – without being concerned with the precise details of what was going on in the other person's mind at the time of the action.

6.3.3. Emotion, mood, and other undue influences on thinking

To explain actions, we often appeal to factors that unduly influence thinking – the way someone's mind works on an occasion – that are not themselves entirely

within what has been called the 'space of reasons': They so to speak bridge the divide between the mental and the physical. They include states such as being drunk, being under the influences of drugs, having a bad cold and being deprived of sleep.[15] They also include emotions like being angry and being jealous, which, although intentional and capable of being grounded in reasons, can still unduly influence thinking. They also include moods, like being depressed, tense, irritable, full of unlocated sexual desire, which are states that are sort of in between states like being drunk and states like being angry: They are intentional, but less specific in their objects than emotion.

Obviously, these states can affect one's ability to engage in certain kinds of theoretical reasoning: Adding up a column of figures or doing philosophy is harder when drunk or feeling irritable, for example. But in practical reasoning, they can also influence thinking by making salient certain considerations over others, and thereby effecting choices – thus, my use of the expression 'undue influence'. 'He proposed marriage because he was drunk' points towards this kind of explanation in terms of undue influence; the action, of proposing marriage, was intentional, but the implication is that he would not have made the proposal if he had not been drunk. This is the familiar territory of weakness of will, accidie, and of action that one later regrets. Being drunk can also explain things done and not intended (his dropping the glass), and it can also explain the unintended manner in which an intentional action is done (his wavering as he headed towards the door), but it is on undue influences on thinking that I will be focusing.

Consider being depressed. Her deciding not to go to work today can be explained by the fact that she was depressed. This thick explanation goes beyond the explanation that appeals to her rationalising occurrent mental states. Perhaps what went on in her mind at the time was the thought that she might not be able to handle this large and important deal, and the thought that if she didn't turn up to work, then someone else would assume the responsibility for it. Our thicker explanation – she was depressed – goes beyond this by pointing towards an explanation of *why* she had those thoughts. These two kinds of explanation are not in competition. That she is depressed is not a rationalising explanation; it is more in the territory of an 'excusing' explanation (with the Austinian qualification that some excusing explanations get you on the hook rather than off it).[16] Of course being depressed can feature as a reason in a belief-desire explanation ('She visited her psychiatrist because she believed that she was depressed'), but in the cases I am interested in, it does not; it just unduly influences reason and choice.[17]

With this last point in mind, another example can be given using emotion. Your shouting at your child for not sitting up straight in his high chair, in spite of your loving disposition towards her, can be explained by the fact that you were angry. Your action can also be given a belief-desire explanation: You wanted her to sit up straight, and you believed that shouting 'Sit up straight!' at her was the best means of getting her to do this. But why did you have these thoughts? Because you were angry, and because shouting at people who do not do what one wants is characteristic of angry behaviour.[18]

Sometimes, the explanation does not directly refer to the individual's mental condition (drunkenness, depression, anger) and its influence on thinking, but instead it refers to the particular situation that the individual is in: implying, but not stating, that being in that situation typically brings about a certain kind of influence on thinking. The literature in social psychology these days is replete with examples, many of which are surprising when first encountered. In the so-called endowment effect, people tend to value goods which they own much more than they expect to value them before they are 'endowed' with them. In one experimental study, participants predicted that they would exchange a school mug once they owned it for $3.73 on average; but after they owned it, their average exchange price was around $6.[19] So, based on knowledge of this kind of causal influence, it is possible for us to explain someone's apparently 'unreasonable' choices or behaviour in situations of this kind outside the laboratory, by reference to the situation – 'He was a "victim" of the endowment effect' – with the implication that in such situations this is what people typically choose to do, and that this particular person is typical in that respect.[20]

Our everyday psychology is replete with action explanations like these. We come, through experience, through literature and the arts, through reading newspapers and history books, through knowledge of experimental psychology, to know what influences various kinds of factors typically have on people.[21]

6.3.4. Narrative-historical explanations

This leads me to my fourth and final kind of thicker explanation, which pulls together the first three, and goes beyond them. We often seek explanations of why someone had a particular motive, or why someone has a particular character or personality trait, or why someone was drunk, depressed or angry. And the explanations that we get are narrative-historical explanations: They locate the motive, the trait, the undue influence on thinking, within a wider nexus, in a way that enables us to explain more deeply why someone did the thing that they did. She acted inconsiderately because she was brought up in a family where considerateness in any form was always taken advantage of. He shouted at his child (in spite of being a loving father) because his job has been under threat and he cannot bear his boss; he is as we say 'taking it out on' his defenceless child. She gave up going to dancing classes because she has post-natal depression and a difficult, neglectful husband.

Narrative explanations, it is generally accepted, are causal explanations.[22] Perhaps, as David Lewis insists, all explanations of events (and thus of actions too) are causal.[23] I am happy to agree to this, and, in fact, it lends some support to my claims about thick action explanations, once we appreciate what Lewis himself emphasises, the 'multiplicity of causes and the complexity of causal histories'; these, he says, 'are obscured when we speak, as we sometimes do, of *the* cause of something'.[24] I want to say something very much like this about the second monopolistic claim about belief-desire explanation of action: It supposes that explanation of action will be sufficient if it refers only to the belief and desire that caused it.

But, except in the special circumstances that I have mentioned, the belief-desire explanation is usually singularly unhelpful or even redundant, true as I accept it is.[25] What we are after, usually, is the thicker explanation, which goes further into the 'countless distinct, converging causal chains', which culminate in the action.[26] Some of these explanations, I have been arguing, are person-specific, pointing, for example, to someone's particular and perhaps unusual mental disposition. Other explanations explain by reference to a typical pattern that people's lives tend to take, from which particular individuals can, of course, diverge. We know that people often do things like angrily shout at their child when they are tense, tired and a bit drunk; and that people often drink, and get tense and tired, when their jobs are under threat. It is all very human and understandable. He might say that he wants the child to sit up straight and believes that shouting at it is the best means of achieving this end. The two explanations are possibly both true (remember again that I do not reject the truth of belief-desire psychology); but the narrative explanation reveals *why* his child's failure to sit up straight made him so angry and have the beliefs and desires that he did have. He might himself be able to give this thicker explanation – although most likely only after the event – and in giving it he would not be giving a belief-desire explanation.

One possible response to the line of argument that I have been advancing would be to say that all these thicker explanations do not really explain *action*, for this is what belief-desire explanations do; rather, they name the causes of the causes of the action; they name the causes of the beliefs and the desires. There is some truth in this, but the response misses the force of the point that, in giving the thicker explanation, one is *thereby* explaining the action in a non-redundant way. We know that he left the ice cream on his plate, and, to a rough approximation, we know the thin explanation, that he did this because he desired not to eat the ice cream, believing it to be strawberry-flavoured, and that he desired chocolate flavour and not strawberry. But does his action reveal a certain inconsiderateness towards his host? Or does it reveal a more general and widespread lack of manners? Or does it reveal a kind of compulsion about always getting what he wants? Or does it reveal a lifetime's loathing of strawberries which can be traced back to, and explained by, that gluttonous day of his sixth birthday?

Answers to questions such as these are very often the concern of our everyday action explanations, going far beyond what is going on in the mind of the person doing the action. And once this is accepted, as I think it should be, we should also come to see that it is a mistake, often made in current philosophy of mind, to focus almost exclusively on trying to solve the mysteries of 'mindreading', in order to explain how we 'gain access' to others' mental states. Belief-desire psychology is only a small part of our everyday psychology.

NOTES

1 *Intention*, Oxford: Blackwell, 1957.

2 See especially Donald Davidson's 'Actions, Reasons, and Causes', together with the other papers in his *Essays on Actions and Events*, Oxford: Clarendon Press, 1980.

[3] For example, Thomas Nagel has claimed that, in respect of some actions, appeal to a belief is sufficient to explain it (*The Possibility of Altruism*, Princeton: Princeton University Press, 1970); Rosalind Hursthouse has claimed that actions expressive of emotion do not involve 'the ascription of a suitable *belief*' ('Arational Actions', *Journal of Philosophy* 88, 1981, pp. 57–68, at pp. 58–59); Michael Stocker has made a similar claim in respect of actions done out of friendship (*Valuing Emotions*, Cambridge: Cambridge University Press, 1996). For disagreement, see Donald Davidson, 'Problems in the explanation of action', in his *Problems of Rationality*, Oxford: Oxford University Press, 2004, pp. 101–116), Michael Smith, 'The Possibility of Philosophy of Action', in J. Bransen ed., *Human Action, Deliberation and Causation*, Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 17–41, and my *The Emotions: A Philosophical Exploration*, Oxford: Clarendon Press, 2000, pp. 125–129.

[4] Davidson, 'Actions, Reasons, and Causes', p. 4; Fodor *Psychosemantics*, Cambridge, MA: MIT Press, 1987, especially Chapter 1; Currie and Ravenscroft, *Recreative Minds*, Oxford: Oxford University Press, 2002, p. 51; Shaun Nichols and Stephen Stich, *Mindreading*, Oxford: Oxford University Press, 2003, p. 4. Nichols and Stich do allow for 'thicker' psychological states, such as *schadenfreude*, although their 'boxological' diagrams have boxes only for beliefs and desires; see p. 14, 31, 40 and 41.

[5] For example, Nichols and Stich say that one of the reasons they 'have opted for "mindreading" is that the association with telepathy infuses the term with an aura of mystique, and we think the capacity to understand minds deserves to be regarded with a certain amount of awe' ibid., p. 2.

[6] It does not just depend on what that rational person desires (although no doubt it does depend on that too); for example, if you were set on showing how resentful you are at being given the wrong flavour of ice cream, it could still make sense to do any one of the things just canvassed (if you were to eat what you are given, you would need to eat it *resentfully* – something easily done).

[7] It is also a weakness, in that it becomes very hard to pin down exactly what rationality is, and what the relation is between some notion of ideal rationality, and the sort of rationality that we need to ascribe to people in order to explain what they do. As Adam Morton puts it, 'Rationality seems both vital and irrelevant to what we actually do' (in *The Importance of Being Understood*, London: Routledge, 2003, p. 40). Shaun Nichols and Stephen Stich are highly critical of Daniel Dennett's interpretive presupposition of rationality; see their *Mindreading*, pp. 142–148.

[8] *Intention*, Sect. 26.

[9] Ibid., Sect. 12.

[10] Nagel, *The Possibility of Altruism*.

[11] I discuss this in some detail in my *On Personality*, London: Routledge, 2005.

[12] For discussion, see my *On Personality*, pp. 8–10, and David Lewis, 'Causal Explanation', in his *Philosophical Papers Volume II*, New York: Oxford University Press, pp. 214–240, at p. 221.

[13] See, for example, Gilbert Harman, 'Moral Philosophy Meets Social Psychology', *Proceedings of the Aristotelian Society* 99, 1999, pp. 315–331, and his 'The Nonexistence of Character Traits', *Proceedings of the Aristotelian Society* 100, 2000, pp. 223–236, and John Doris, *Lack of Character: Personality and Moral Behaviour*, Cambridge: Cambridge University Press, 2002.

[14] See my *On Personality*, Chapter 3.

[15] These kinds of influences on thinking are discussed by Jane Heal in 'Replication and functionalism', in M. Davies and T. Stone, eds, *Folk Psychology: The Theory of Mind Debate*, Oxford: Blackwell, 1995, pp. 45–59. I discuss them in *The Emotions: A Philosophical Exploration*, pp. 167–175.

[16] J.L. Austin, 'A Plea For Excuses,' reprinted in his *Philosophical Papers*, 3rd edition, Oxford: Oxford University Press, 1979, pp. 175–204.

[17] Christine Korsgaard discusses related issues in her 'Skepticism About Practical Reasoning', *Journal of Philosophy* 83, 1986, pp. 5–25.

[18] I discuss examples like this one, and the notion of how emotion can 'skew the epistemic landscape', in 'Emotion, Reason, and Virtue', in P. Cruse and D. Evans, eds, *Emotion, Evolution and Rationality*, Oxford: Oxford University Press, 2004, pp. 247–266.

[19] Nichols and Stich, *Mindreading*, p. 138.

[20] There is something of an asymmetry here between first-personal and third-personal explanations, and where the action being explained is actually going on at the time that the explanation is being given. He is shouting at his child because he is angry, but I am shouting at my child because he *must* learn to

sit up straight. But when we are looking back on our past actions, this asymmetry tends to drop away (I'm sorry, it wasn't fair, I was angry). I discuss this in relation to planning in 'Imagination and the Distorting Power of Emotion', *Journal of Consciousness Studies* 12, 2005, pp. 130–142.

[21] Donald Davidson makes what I think is a similar suggestion in 'Hempel on Explaining Action', in his *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, pp. 261–275. He writes of the importance of 'general knowledge of the nature of agents', but insists, as I do, that 'such knowledge is not used in giving reason explanations' (p. 272). Davidson's position differs from mine, however, in part because he also insists that reason explanations are 'satisfying and informative', and that they are made 'more valuable by letting us fit them into a larger scheme' (p. 272, 273). Perhaps this is just a matter of emphasis; I am not sure.

[22] See, for example, Noel Carroll, 'On the Narrative Connection', in his *Beyond Aesthetics*, New York: Cambridge University Press, 2001, pp. 118–133, and David Velleman, 'Narrative Explanation', *Philosophical Review* 112, 2003, pp. 1–25.

[23] See his 'Causal Explanation'.

[24] Ibid., p. 215. Lewis' example is of the car crash: 'we have the icy road, the bald tire, the drunk driver, the blind corner, the approaching car, and more'; 'The roots in childhood of our driver's reckless disposition, for example, are part of the causal chain via his drunkenness, and also are part of other chains via his bald tire'. Ibid., p. 214 and 215.

[25] Compare the true statement that the cause of the crash was the car leaving the road.

[26] Ibid., p. 214.

DANIEL D. HUTTO

7. FOLK PSYCHOLOGY WITHOUT THEORY OR SIMULATION

> If commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species
>
> Fodor (1987, p. xii)

7.1. FOLK PSYCHOLOGY STRICTO SENSU

Folk Psychology is dead. Long live Folk Psychology!! This could be the motto of many of the papers in this volume. I too endorse the need to reform the standard assumptions about the function, scope and basis of our capacity to understand others in terms of what might be called – accurately, if rather cumbersomely – propositional-attitude belief-desire psychology. Yet, I stop short of proposing a successor. I take it to be a datum that certain populations of psychologically normal, adult humans do, as a matter of fact, make sense of intentional action by appeal to reasons. In speaking of 'reasons' I mean precisely what philosophers have long understood to be at the heart of discrete episodes of means-end practical reasoning – processes that result in intentions to act (see, for example, Goldie, this volume).

It is a commonplace that we make sense of actions in such terms (I say more about the reference of this 'we' in a moment). Sometimes we act for one reason and not another, decisively; though I am happy to grant that this may be a less frequent occurrence than is commonly supposed. Equally, determining for which reason an action was performed may be extremely difficult, even for an action's author. Nevertheless, if an action is done for a reason it must be possible to explicate it, minimally, by appeal to a *particular* belief/desire pairing. As the essential components of reasons, these psychological attitudes must each stand in relation to distinct propositional contents. These contents connect the attitudes in virtue of their having certain overlapping elements. Maximally, reasons can be explicated – in more or less refined ways – by detailing yet other propositional attitudes (hopes, fears) and other more basic kinds of perceptions and emotions. Indeed, a complete explication would also need to give details of the person's character, situation and history – in short, his or her 'story' (see also Gallagher, Goldie this volume). To keep things straight, when I talk of folk psychology I am *solely* concerned with the practice of predicting, explaining and explicating intentional actions by appeal to reasons understood in this way. At a bare minimum, folk psychology *stricto sensu* is belief/desire propositional attitude psychology.[1]

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 115–135.

Without doubt the so-called friends of folk psychology have overstated and misunderstood its role in social cognition; typically they see it as (1) more basic and (2) far more pervasive than it actually is. With respect to the question of primacy, folk psychology is not *fundamental* to social engagements; not even exclusively human ones. We have many other, more basic non-folk psychological means of engaging with one another socially and coordinating our interactions – these involve only end-directed intentional and not propositional attitudes. Our primary embodied modes of responding, by my lights, do not even involve the manipulation of representations by inferential operations (let alone representations of propositional attitudes). Nor do such engagements result in predictions or explanations, understood as couched in sub-personal propositions. In such cases we get by with script-like patterns of recognition-response (some more flexible and complex than others): These are initiated and guided indexically and iconically by the expressive behaviour of others. These sorts of abilities – and not a capacity for 'mindreading' – best explain the embodied expectations of non-verbal creatures. In 'normal' contexts these modes of response are not only quicker, they are highly effective ways of interacting and navigating social dynamics (see Hutto 2006a, b, 2007a). In saying this, I fully support the conclusions of several contributors to this volume (Andrews, Hobson, Gallagher and Zahavi).

With respect to scope, I claim that folk psychology stricto sensu is the unique provence of certain linguistically competent human beings (*Homo sapiens sapiens*) – at least those who do not suffer from extreme autism.[2] By my lights, reason-based understandings are not used by our close living cousins, the chimpanzees, nor were they used by our ancient ancestors who hailed from the Pleistocene (with the possible exception of the early humans of Upper Palaeolithic – and this, as will soon be clear, depends on the nature of the discursive practices they had). Indeed, even with respect to contemporary human populations folk psychology may not be a universal good.

It is not a given that all cultures engage in the practice of understanding actions in terms of reasons. Evidence gleaned from the handful of cross-cultural studies that have been conducted suggests that neither an understanding of folk psychology nor a command of the concept of belief come automatically or in equal measure to all unimpaired members of our species. Tests conducted with children from several non-Western cultures reveal that they do not employ the folk psychological schema as readily or with the same proficiency as Westerners do (Vinden 1996, 1999; Lillard 1997, 1998).[3] This does not show that the populaces of these cultures *never* make sense of one another in terms of reasons for acting (and it certainly does not show that they are incapable of doing so) but it does raise questions.[4] Looking carefully at the practices of other cultures, one finds heterogeneity in the explanatory tendencies and methods used in understanding action[5] – in some parts of the globe greater emphasis is placed, for example, on the situational, trait-based and even supernatural explanations.[6] The evidence, such as it is, should make us cautious of simply assuming that *all* human cultures share an understanding of belief/desire psychology. And it will become evident momentarily why we must be especially cautious in

making this assumption about those cultures whose intellectual history and narrative practices differ significantly from those of the West (see also Kusch, this volume).

As discussed in the introduction of this collection, for several decades the dominant view has been that folk psychology is either a kind of low-level theory about the propositional attitudes or a simulative ability involving their direct manipulation (or a mix of both). In place of these options, I propose that we should understand folk psychology as – in essence – a particular kind of narrative practice: This would best account for its origins and various applications. According to the narrative practice hypothesis (NPH) it is through direct encounters with stories about reasons for acting, those supplied by responsive caregivers in interactive contexts, that children become familiar with (1) the core structure of folk psychology and (2) the norm-governed possibilities for wielding it in practice (i.e. learning both *how* and *when* to apply it). A distinct kind of narrative practice, one involving a particular kind of story, plays the central role in readying us to understand intentional actions folk psychologically. My proposal is that encounters with stories about those who act for reasons are what acquaint us with the *forms* and *norms* of folk psychology.

I give full details of this conjecture and its implications elsewhere (Hutto 2007a, b see also Gallagher 2006, Gallagher and Hutto 2007). My purpose in this paper is merely to spell out just *how* the NPH, if true, undercuts any need to appeal to either theory or simulation when it comes to explaining the *basis* of folk psychological understanding: These heuristics do not come into play other than in those cases in which the framework is used to speculate about how another might act or why another *may* have acted. To add appropriate force to this observation, I will first say something about why we should reject the widely held assumption that the *primary* business of folk psychology is to provide third-personal predictions and explanations. I then go on to demonstrate how the NPH can explain (1) the structural features of folk psychology and account for (2) its staged acquisition without buying into the idea that it is a theory, or that it is acquired by means of constructing one. This should eliminate some of the more popular reasons for believing that folk psychology *must* be a kind of theory. In the concluding postscript, I acknowledge that we need more than the folk psychological frameworks in order to understand reasons, but I deny that this something more takes the form of a theory about beliefs and desires or simulative procedures for manipulating them. For example, I claim it rests in part on a capacity for co-cognition, inter alia, since that ability is necessary for understanding another's thoughts. Nevertheless, I deny that co-cognition equates to simulation proper or that it plays anything more than a supporting role in enabling us to understand reasons for action.

7.2. THE PRIMACY OF SECOND PERSON APPLICATIONS

Folk psychological narratives function as 'normalising' explanations, allowing us to cope with 'unusual' or 'eccentric' actions, where possible, by putting them in context – this either helps us to see why they fall within the fold of the normal or

it may extend the bounds of what is regarded as normal (failing this it will not be possible to make full sense of why the action was performed).[7]

The restorative accounts that such narratives provide need not issue from the person *seeking* the explanation. The greatest chance of obtaining a successful explanation – of deciding for which reason an action was performed – depends on the authors of actions identifying and explicating their reason, for themselves (Bruner 1990). Other conditions must hold too – i.e. the person must not be confabulating; he or she must not be engaging in post-hoc rationalisation; he or she must not be self-deceived, etc. Nevertheless, by way of comparison, asking the other for their reasons is *vastly* more reliable than trying to determine why they *in fact* acted as they did from the distance of a third-party spectator. Indeed, in such attempts the accuracy of an explanation is likely to be inversely proportional to the need to seek it at all (Hutto 2004). It is only in second-personal contexts that we confidently obtain *true* folk psychological explanations, by and large, as opposed to speculating about merely *possible* ones. When in doubt it is best to get one's explanations from the 'horse's mouth', as it were. Even though in some cases we will have legitimate reasons to doubt the other's word, the 'explanations' that we generate on their behalf rarely rise above the status of mere supposition (at least in those cases where there is any interesting question about their reasons for acting in the first place).

The stories others tell about their reasons are typically delivered, and indeed, fashioned in the course of on-line interactive dialogue and conversation – dialogue of the sort that is, with luck, sensitive to questioner's precise explanatory needs and requirements. The nature of such engagements is complex and deserves more attention than it has received to date (see Stawarska, this volume). That is not my focus here. The crucial point to recognise is that it is these second-person deliveries – these *narrations* – that do the heavy lifting in enabling us to understand and make sense of others with confidence. I call narratives of the kind that explicate actions in terms of reasons (as restrictively defined above) *folk psychological narratives*.[8] Providing these is the *primary* work of folk psychology stricto sensu. Thus, by my lights, folk psychology in action *is* – at bottom – a distinctive kind of narrative practice.[9] In a derivative way, its framework can be used for third-personal speculation as well – as in those cases in which we wonder why another *may* have acted on a particular occasion. This may be carried off by various means, but I want to stress that it is always a parasitic and peripheral business. I now turn to the question of the origin of the folk psychological framework itself, which I take to have emerged from encounters with folk psychological narratives.

The preceding observations about the importance of second-personal narrations concern not only our contemporary dealings, but also those of our ancient ancestors. I claim that the folk psychological framework not only has its *primary application* in second-personal engagements, it *emerged (and emerges)* from these as well. The above observations matter when it comes to assessing the standard conjectures about the ultimate origins of our folk psychological abilities, for they imply that third-personal mindreading – of the sort involving the attribution of the core propositional

attitudes – simply is not a reliable basis for sophisticated reason-based understanding of the actions of others. If so, this ought to cast serious doubt on the idea that the folk psychological framework was originally *put in place* because it served this purpose. And when seriously examined, by giving attention to the kinds of pre-linguistic social interactions in which our ancient ancestors are likely to have been engaged, the received wisdom about the origins of folk psychology looks deeply implausible. The relatively sophisticated activities of our forerunners are best understood in terms of a range of interactive, imaginative and mimetic capacities, none of which presuppose any 'folk psychological' understanding on their part at all. Indeed, when abductively compared with a different hypothesis about the imaginative and mimetic abilities of hominids, 'mindreading' proposals look exceedingly weak. As I argue elsewhere, *pace* Mithen, metarepresentational theory of mind abilities are not needed to explain hominid (1) tool-making, (2) social cohesion, or even (3) basic interpretative and language learning abilities (see Hutto 2007c; Mithen 2000a).

In this light, we would do well to rethink the role that such 'explanations' have in our lives, since their function is clearly *not* primarily to enable us to generate third-party speculations (not even in the form of gossip) about why others *may* have acted thus and so. This in no way diminishes – indeed it may well enhance – our understanding of the importance of folk psychology and its place in our activities and practices (see Andrews, Knobe this volume).

In thinking about the prehistoric origins of folk psychology, it is important to remember that the practice involves more than just the wielding of the mentalistic concepts of the attitudes; it requires that one has an appropriate way of characterising the propositional contents that serve to describe *what* it is that the other believes and desires. It follows that any would-be folk psychologist must be capable of 'representing representations' – more precisely, representing representations that have propositional forms and contents. We can be sure of this because folk psychologists trade in *reasons* – minimally, belief/desire pairs – not just isolated thoughts or desires. When understanding action as performed for a reason it is not enough to appeal to a lonesome propositional attitude – the ascription of its relevant partner or partners is also required. Furthermore, these attitudes and their partners, so described, must be understood as directed at representations with appropriate logical forms – for it is in virtue of the formal properties of such representations that local inferences are made, yielding intentions to act.

By implication, if one were restricted to understanding action using only holophrastic representations – those with no proper internal parts – it would be impossible to represent inference-based thinking at all (including basic practical reasoning, decision-making or planning). Such a creature could not understand or attribute *reasons* since doing so entails having the ability to represent a complex 'state of mind' in which multiple psychological attitudes (i.e. beliefs, desires, hopes, etc.) are directed at interlocking propositional contents – those, in turn, are linked by some overlapping internal elements. The propositional contents in question therefore must be composed of distinguishable, recurring semantic elements.

For this reason, in order to model a mind capable of representing representations of the appropriate kind requires a commitment to *some* kind of sententialism. Sententialists claim that even the *having* of propositional attitudes should be understood as instantiating a three-place relation in which thinkers stand in relation to sentences – adopting various psychological attitudes towards them – and in turn, the contents expressed by these sentences themselves 'picture' some specific states of affairs. On the traditional analysis, the meaningful parts of these sentences refer to worldly objects and the expressed content of the sentence as a whole will be true if the world obliges – i.e. if the relevant state of affairs that the sentence describes, as picked out disquotationally, obtains (otherwise it is false). In this respect, sentences have just the right kind of semantic properties of reference and truth required to be the appropriate *relata* of the attitudes. Sentences are purpose built for the required work. They have internal logical forms and syntax, and as such they are tailor-made for explaining the computational, inferential features that propositional attitudes exhibit, assuming as we must that it is the propositions of propositional attitudes that do all the interesting logical work.[10] Being spatio-temporal particulars, and not abstract objects, sentences have the added virtue of being the kinds of things to which one might be causally related.[11]

One way of accommodating sententialism – the one I favour – is to suppose that propositional attitude-based reasoning and its representation is conducted using the public vehicles of natural language (or surrogates derived from them).[12] Accordingly, the structures provided by context-invariant linguistic symbols are the very basis for inferentially based modes of reasoning. On this account, the symbol systems of natural languages, and the compositional semantics they provide, will have made bona fide inferential thinking possible for the first time. Such external formalisms are the necessary ingredients for genuinely logical reasoning (Clark 1989, pp. 132–133). Although it seems that our non-linguistic hominid ancestors must have been capable of a *kind* of consequent-sensitive thinking, it would not have been of the strictly logical variety. Most likely, it would have been based in the manipulation of images supplied by the recreative imagination (Hutto 2007a). Only on acquiring complex natural language would our forerunners have come by the structures needed for unrestricted logical thinking of the sort that involves the manipulation of discrete propositional forms and their sub-components (Frankish 2004, ch. 6 and 7). For convenience, I follow Davies in labelling this the 'thinking in natural language' hypothesis (TNLH) (Davies 1998, p. 226). Its central idea is captured in the memorable slogan: "The language of thought is natural language" (Frankish 2004, p. 197).[13]

In explicating the TNLH it helps to call on Frankish's handy distinction between the basic mind and supermind (Frankish 2004).[14] As I characterise this divide, basic non-linguistic cognition, of any variety, should not be understood in terms of the use of propositional representations at all.[15] In direct contrast, supermental thinking takes the form of making active, conscious commitments – the adoption and maintenance of premising policies – policies concerning the use of propositions as premises in truth-seeking and means-end reasoning. The supermind

is thus a linguistically based, soft-wired virtual machine that imposes structural regularities on our underlying mental hardware, substantively altering its habits and dispositions.[16]

Against this background, when it comes to explaining the origins of the folk psychological framework I take a leaf out of Sellars' book; both in making clear the limit of my ambitions but also in offering much the same sort of account he offers about how our understanding of 'inner episodes' of thought was first forged. Sellars imagines a mythical Jones who models thoughts on overt speech acts. In a structurally similar way, I hold that reasons – minimally, logically interlaced belief/desire pairings – would have been initially modelled on overt, temporally extended, public accounts of episodes of practical reasoning – those involving the explicit manipulation of logically complex symbols with recurring internal parts – i.e. natural language sentences. For it is towards these (or their 'inner speech' proxies) that reasoners' attitudes are directed.

In the giving of such accounts acts of reasoning would have been put on exhibit, taking the form of third-personal representations. This would have taken place in second-personal dialogical contexts, within which these representations would have been complex objects of joint attention. Crucially, the arrival of such folk psychological narratives must have post-dated that of linguistically based superminds. This is so for two reasons: (1) the capacity to reason practically requires facility with a compositional language (assuming the TNLH is true) and (2) the capacity to *describe* the moves made in such reasoning also requires facility with a compositional language. If we accept the standard dates of the origin of such language, circa 35–40,000 years ago, our folk psychological abilities cannot be explained by a more ancient endowment.

Following Sellars' line of reasoning, I propose that the very first psychological narratives, on this view, would have been related in public by practical reasoners themselves. The authors of certain actions would have given accounts of the plans they constructed based on propositional beliefs and desires, at least. By listening to such narrations, the framework of folk psychology would have been discerned again and again, eventually becoming available for other potential uses. Third-personal folk psychological speculation would have followed on the heels of second-person folk psychological narration – and it looks likely that both of these practices were rather late developments in the socio-cultural history of our species.

Allying myself with Sellars in this way may appear to be an odd move – after all, he is frequently presented as an arch theory theorist, indeed possibly even the first of the kind. It is not uncommon to hear that "Early formulations of the notion of folk psychology stressed the idea that folk psychology is an *explanatory theory*. This is much to the fore, for example, in Sellars' influential mythical account of how folk psychology might have emerged" (Bermúdez 2003, p. 47, emphasis added). But this reads much too much into the parallel that Sellars drew between the *context* in which non-observational 'inner episodes of thought' were first constructed and the *context* in which theoretical posits are constructed. Sellars *only* ever claimed that his "story helps us to understand that concepts pertaining to such inner episodes

are primarily and essentially *intersubjective*, as intersubjective as the concept of a positron, and that the reporting role of these concepts – the fact that each of us has a privileged access to his thoughts – constitutes a dimension of the use of these concepts which is *built on* and presupposes this *intersubjective* status" (Sellars 1956/1997, p. 107). I too want to stress the intersubjective basis of our understanding of reasons, but clearly saying only this does not commit me, or anyone, to the idea that such an understanding is theory-based, theory-like or formed as a product of theorising. To think otherwise, would be to be confused about the scope and nature of the claim that both mentalistic concepts and theoretical constructs have an intersubjective basis – they need be alike in no other respect.

The true genius Jones, of Sellars' infamous myth, as I have remixed it, was not a theoretical mastermind after all but merely an attentive listener to the stories of his fellow practical reasoners (cf. Sellars 1956/1997, pp. 102–103). His great innovation would have been to adopt the folk psychological framework, as revealed by second-personal dialogues of the specified type, putting it to (a rather degraded) work in third-personal speculation about why others *may* have acted thus and so on particular occasions. If this is right, we have no legitimate grounds for calling the core folk psychology framework, even so deployed, a 'theory of mind'.

Widespread, now well-established, social practices involving distinctive kinds of narratives play the role that many have postulated *must* be played by inherited mechanisms. It should therefore be possible to explain how budding folk psychologists come by a practical grasp of the core folk psychological concepts (as well as the ability to structurally represent *how* these relate in the context of reason explanations, schematically as it were), without postulating any hard-wired mechanisms that already contain this information in the form of a theory of mind or by postulating mechanisms that allow the construction of one. This is the burden of Section 7.3.

7.3. THE STRUCTURE AND STAGED ACQUISITION OF FOLK PSYCHOLOGY

He left the party because he believed the host had insulted him. She will head for the cabin in the woods because she wants peace and quiet. These are typical examples of reason explanations; one backward looking and the other future facing. Both imply more than they say. To leave a party because of a suspected insult suggests that one desires not to be insulted, or at least that the desire to avoid insult is stronger than that for some other good on offer. Similarly, to seek tranquillity in an isolated cabin implies that one believes that it can be found there, or at least more so than elsewhere. Despite the fact that these explications of the reasons for taking these actions are woefully under-described, they demonstrate that a major aspect of making sense of action in terms of reasons rests on a quiet understanding of the way propositional attitudes interrelate.

How do we come by this during childhood? The capacity to understand reasons for acting rests on a complex series of foundations – it is a multi-layered ability. Long

before acquiring it, those children who develop normally are able to navigate the social world using embodied skills, which require no understanding of reasons for action whatsoever. With a growing command of language, they acquire a practical grasp of the different kinds of propositional attitudes, individually. I will say more about this in a moment but for now it is only important to note that children come into the possession of all the pieces needed for playing the understanding-action-in-terms-of-reasons game, before they can actually play it.

Having all the pieces is necessary but not sufficient for such play. And it is well known that children make *propositional attitude* ascriptions before they are able to explicate, explain or predict actions in terms of reasons. At around two years of age, children are in secure possession of "an early intentional understanding of persons having internal goals and wants that differ from person to person" (Wellman and Phillips 2001, p. 130; Bartsch and Wellman 1995, ch. 4). Their understanding can be rather sophisticated: They refer to unfulfilled and future desires, exhibiting some fluency with counterfactuals.

But just as such *desire* ascriptions on their own should not be confused with *belief* ascriptions, so too *belief* ascriptions on their own should not be confused with *reason* ascriptions. Thus, obtaining an understanding of metarepresentational thinking – getting a handle on the concept of belief – is not the final step in acquiring folk psychology abilities.[17] One can ascribe beliefs using a simple inference rule of the following sort, if X says that P then X thinks that P. Knowing that X thinks that P is useful for certain coordinating purposes – for example, it serves as the basis for predicting what else X might think (on the assumption that X observes standard norms of rationality). I say more about this in the postscript discussion.

An example of Millikan's serves to underline the main issue: for many coordinating purposes it can be enough to know simply that John *likes* or *wants*, say, yoghurt (Millikan 2004, pp. 21–22). Young children are certainly capable of noting this sort of thing – this is what enables them to make certain low-level, inductively driven predictions. But doing so does not equate to understanding John's action in terms of his reasons; for that, more is needed. In particular, the child would have understood that John's action issued from a 'complex state of mind, one with a specific implicit structure: this structure is what is made explicit, at least partly so, when one says John is eating yoghurt for breakfast *because* he believes it will make him healthy – implying also that this is something he wants.

Proficiency in making isolated propositional attitude ascriptions – attributing certain goals, desires, thoughts and beliefs – is not the same as knowing *how* to combine these in order to understand actions in terms of reasons. This stronger condition must be satisfied if one is to be a folk psychologist. Not only must children have an understanding of the core propositional attitudes – belief and desire – they must also learn how these interrelate with one another and other standard players in psychological dramas (for details see Hutto 2007b). The crowning requirement for acquiring folk psychology is that children must master the norms that detail the interplay between the various propositional attitudes – attitudes of which they already have a prior, discrete practical grasp. In essence, they must be familiar not only with all

the elements in play but also with how they can be combined in appropriate ways. What is missing is therefore not another ingredient in the folk psychological cake, but instructions on how to mix the existing ingredients properly to make many such cakes. But if such instructions are not built-in mechanically, how do children acquire them?

Folk psychology is a complex skill, the full mastery of which only comes over time. Children gain it if they have the right inherited capacities and if they are appropriately supported by their elders while engaging in specific kinds of story-telling practices, i.e. folk psychological narrative ones. Mirroring my Sellarsian proposal about the ontogenesis of the framework of folk psychology in our pre-history, stories detailing the reasons for which protagonists act serve as exemplars that introduce the very same framework to individual children in ontogeny. It is through repeated, interactively guided encounters with such stories that the *forms and norms* of folk psychology are revealed. During this process, children call on a range of imaginative and embodied capacities and are directed in important ways by their carers. Crucially, the relations that hold between mentalistic concepts and the normal contexts in which they operate are laid bare, allowing children to become familiar with the folk psychological schema and the norms governing its practical application. Only certain types of narratives enable this: *folk psychological* narratives about reasons.

There are two senses of 'narrative' to distinguish here – narrative in the sense of the third-personal object of focus – the folk psychological narrative itself – and acts of narration, the second-personal interactions that constitute the story-tellings through which children are introduced to such. Understanding both, without conflation, is crucial for understanding the NPH.

As an object of co-attention, the narrative or story itself might be a spontaneous production, an autobiographical account or a bit of gossip, or the retelling of a set text (usually taking the form of an established cultural artefact, of which there are, typically, multiple versions). Here's one such narrative (as yet I have no data on how many of these children encounter in the normal course of their development: I leave it to the reader to speculate about this).

> Little Red Riding Hood *learns* from the woodcutter that her grandmother is sick. She *wants* to make her grandmother feel better [she is a nice, caring child], and she *thinks* that a basket of treats will help, so she brings such a basket through the woods to her grandmother's house [beliefs and desires lead to actions]. When she arrives there, she *sees* the wolf in her grandmother's bed, but she *falsely believes* that the wolf is her grandmother [appearances can be deceiving]. When she *realizes* it is a wolf, she is *frightened* and runs away, because she *knows* wolves can hurt people. The wolf, who indeed *wants* to eat her, leaps out of the bed and runs after her trying to catch her (Lillard 1997, p. 268, emphases mine).

Tales of this sort are the most reliable means of exhibiting how the core elements needed to understand reasons work together. Such stories include, inter alia, illustrations of the relation of ends and means, the way projects stack, how one's purposes can be at odds with those of others, and so on. But, crucially, they have *precisely* the right *form* and *content* for showing how the core propositional attitudes interact in reasoning, revealing their proper inferential relations and roles.[18] This provides the necessary framework for reflectively applying mentalistic concepts when under-

standing the unexpected intentional actions of others (and understanding one's own) in terms of reasons.

A major virtue of the NPH is that it can uphold Lewis's observation that the 'meaning' of mental predicates, at least when they cooperate in the context of reason explanations, is determined by their role in a structured framework. The caveat matters since I hold that children have a practical grasp of the individual propositional attitudes independently of their understanding 'reasons'. But although Lewis may have been right about what is required for understanding reasons, we should not conclude that the folk psychological framework is a theoretical one. We cannot conclude that even if having a framework structure should turn out to be an essential feature of theories – one that determines the 'meaning' of their constructs as well. It is worth recalling that Lewis originally illustrated his claim about the meaning constituting properties of such structures by appealing to the putative 'suspect theory' inherent in Cluedo (Lewis 1970, 1978). But as a matter of fact, Cluedo doesn't have an inherent 'theory'. It has a set of rules which one must master in order to play the game. But these rules are not theoretically grounded, they are conventional – to learn them is to learn the rules of a certain established social practice. I am not suggesting that the narrative practice of explaining ourselves by citing reasons is a rule-bound game like Cleudo. Rather I simply want to emphasise this: Just having an inferential structure does not make something into a theory.

So, we might agree with Lewis that our understanding of the mentalistic predicates that comprise 'reasons for action' are best understood in the way we understand other theoretically embedded vocabularies (i.e. talk of electrons, atoms and gravity). In both cases the meaning of such concepts may be fixed, at least in part, by structural links. But again, this does not make such concepts *theoretical*. At best, they are similar in this one respect with theoretical terms. It is important to realise that this is *only* because theoretical concepts are a subset of the kind of concepts that gain their meaning holistically: it is not because the concept of 'reasons for acting' is a theoretical construct (the same holds for other mental predicates).

Finally, the NPH sits easily with the fact that the folk psychological components needed for playing the folk psychological game are acquired separately and in stages. But to explain this we do not need to understand them as the products of dynamic theory-building. Perhaps the major attraction of the 'scientific theory theory', especially for developmental psychologists, has been that it looks uniquely well suited to explain how mentalistic concepts are acquired (indeed, forged) in a punctuated way. Thus, a burden of rival 'modular' accounts that postulate in-built theory of mind mechanisms has been to demonstrate that they too can accommodate this seeming fact, either by explaining it or explaining it away. Thus:

> The developmental evidence suggests that children *construct* a coherent, abstract account of the mind which enables them to explain and predict psychological phenomena. Although the theory is implicit rather than explicit, this kind of cognitive structure appears to share many features with a scientific theory. Children's theories of the mind postulate unobserved entities (beliefs and desires) and laws connecting them, such as the practical syllogism. Their theories allow prediction, and they change (eventually) as a result of falsifying evidence (Gopnik 1993, p. 333, emphasis added).[19]

Yet, it is possible to understand how children acquire the relevant concepts in stages by concentrating on their growing range of *practical* – not their imagined theoretical – abilities. Making this our focus, we can interrogate how these abilities develop and their basis, i.e. what underwrites and engenders them. When it comes to understanding the acquisition of 'mental concepts', on the assumption that their general character is already known to us, this type of investigation will yield a unique ability profile for each predicate – detailing the necessary prerequisites for its acquisition, in terms of:

1. More basic abilities/capacities (whether non-conceptual or 'conceptual')
2. Scaffolding supports (e.g. cognitive tools that extend the possibilities for interaction such as linguistic constructions) and
3. Engendering or enabling socio-cultural practices.

Children gain a practical grasp of core mentalistic attitudes in piecemeal fashion. Not only does their grip on these tighten over time, the very nature of what they have hold of changes. This is a multi-staged process. Long before acquiring a practical grasp of the propositional attitudes, unimpaired children are able to navigate the social world using a range of embodied skills, interacting with others in ways which require no understanding of belief or reasons for action whatsoever. With a growing command of language they are able to make use of syntactical constructions, with embedded complement clauses which are new objects of attention and co-attention. In the first instance, this extends their understanding of the possible objects of desire. Some time later, normally about 6 months or so, after exposure to another enabling practice – that of partaking in early conversations in which participants give expression to divergent cognitive takes on worldly offerings – children get a handle on a new kind of attitude, that of belief (and by implication, false belief) (Harris 1996). This requires the exercise of certain of their recreative imaginative abilities, specifically that of visual perspective shifting – only it is employed in a novel context, that of discursive conversation where it is asked to manipulate complex linguistic objects as opposed to non-propositional perceptions (Currie and Ravenscroft 2003; Prinz 2002). Thus around the ages of 3 and 4, children – at least those with the relevant interactive and imaginative abilities and who have taken advantage of the right developmental opportunities – acquire basic capacities to attribute propositional attitudes and to make limited predictions based on such ascriptions. In line with 'supermentalism', it follows that children only come to have propositional attitudes and the ability to represent them after they have mastered certain linguistic complexities.

This is only the briefest thumbnail sketch of the stages and processes by which children come by the components of folk psychology. Due to pressures of space, I cannot go in more detail here (I do so elsewhere, Hutto 2006c, 2007a). But even this skeletal account suffices to demonstrate that we do not need to postulate a theory of mind of any kind in order to account for the early stages through which our understanding of the mental develops.

For all the reasons cited above (and many others not discussed in this essay), I hold that using the label 'theory of mind' as byword for the practice of understanding

intentional actions in terms of reasons – even when no explanatory proposal is attached – is vastly misleading. Given the bad effects it has had and continues to have on the imaginations of many philosophers, psychologists and other researchers it should be completely avoided.

7.3.1. Postscript: the supporting role of co-cognition

I admit that it may be that when we speculate about *possible* reasons for action we must use a low-grade theory or some kind of simulative heuristic or a mix of both. However, it is a major mistake to focus on these peripheral uses of the framework of folk psychology, treating them as if they told us about what lies at the heart of the practice. There is no place for 'theories', 'theorising' or 'simulation' when it comes to understanding the *primary* basis of that ability. Let me be clear, that by 'simulation' I specifically mean these versions that propose that we understand reasons for action by the direct manipulation of the propositional attitudes – at least, beliefs and desires. For I fully accept that when we use the framework of folk psychology in understanding others we must also call on a range of embodied, imaginative abilities that involve non-propositional varieties of 'simulation' (for details see Hutto 2006b, c). Although the NPH tells us how we become acquainted with the forms and norms of folk psychology in a way that shows how we overcome the 'folk psychological' variant of the frame problem, it still only provides *part* of the story of how folk psychology is practically applied (Hutto 2007b). It is thus the tip of a much larger iceberg.

One crucially important background capacity is that of co-cognition – for it is implicated in the digestion of accounts of actions that are done for reasons. But this does not presuppose any 'theory of mind' or 'simulative' abilities proper. I will explain the role co-cognition needs to play before returning to this point.

Consider what is involved in understanding McX's explanation that he reached for the glass of water *because* he was thirsty (to avoid any accusations that this is a 'canned' example, let us suppose that there was an equally good alternative explanation in the air as to why McX might have done this in the circumstances). If his answer is to dispel my curiosity, I must know what anyone can be expected to know about the relevant properties of water. Yet, since the content of any thought is constrained by the content of other thoughts, knowing such things seems to entail a standing capacity to somehow work through an impossibly large sum of inferences (potentially infinite in number). Spelling out what 'anyone can be expected to know about a particular topic' would require explicitly stating all the possible inferential liaisons. In sum, this would constitute a description of the whole of our commonsense knowledge on all topics, on the assumption that it could be laid out in the form of a series of rules and representations.

Thus when it comes to understanding reasons, apart from calling on the core principles of folk psychology, one would be calling on knowledge of an indefinite number of additional principles – principles detailing both commonsense and specialised knowledge of every possible domain of thought. The mere fact that any such 'theory' would require an infinite number of principles is enough to cast its

possibility into doubt.[20] Yet even if we were prepared to countenance its possible existence, questions would loom large about how anyone could possibly wield such principles sensitively in real-time practical applications. For only a small subset of the possible inferences would ever matter in any given case. So, in making sense of McX's answer I somehow just know which ones are the relevant ones on which to focus. There is no formula for achieving this.

How would a theory or theory-driven mechanism determine such things? Deciding which thoughts ought to *go together*, as relevant to a particular judgement made in specific circumstances, is not something that can be specified in advance or once and for all: It is deeply context-sensitive. Judgements of this sort need to be formed on the spot – they are a posteriori. There is simply no algorithm, however complex, that would enable us to anticipate such possibilities. This feature of central cognition is, by its very nature, too unconstrained to be explained computationally. This conclusion is unavoidable if we observe – as we ought – that the non-demonstrative inferences of central cognition are holistic. Indeed, Fodor who has done the most to highlight this difficulty has gloomily pronounced that if a great deal of cognition really were holistic in this way then cognitive science has seen the harbinger of doom (Fodor 1983, part IV; Fodor 2000, ch. 2).[21]

But such holism is only threatening to those who think that in making sense of the thoughts of others we must be operating with a tractable theory of relevance. In response to this problem, which we can call 'Heal's challenge', theory theorists have wisely confirmed the modesty of their position, clarifying the true scope of their commitments (Heal 1998a, b). They have been quick to concede that their proposals *only* concern the core principles of the theory of mind, what I have been calling the folk psychological framework. And they acknowledge that it "is all very well as a framework, but it plainly needs to be supplemented in some way if one is to be able to provide fine-grained intentionalistic predictions and explanations" (Carruthers 1996a, b, p. 24; see also Nichols and Stich 2003, pp. 86, 104).

Thus even hardcore theory theorists typically respond to Heal's challenge by adopting her own account of thought replication, according to which we are able to understand and decide what another is likely to think about any given topic by co-cognising with them. In doing so, we use our own thoughts and all their standard implications as initial guides to the thinking of others, making interpretative adjustments as necessary (see Hutto 1999b, ch. 5). Essentially, co-cognition involves replicating the target's deliberative processes, using one's own thoughts to fuel this activity – thus no principles are involved (cf. Heal 1998b, p. 491).[22] To achieve an understanding of what another *should* infer about a given topic, we need only call on our own first-order commonsense knowledge about the world.[23] For example, provided that X has reliable information about Y's initial thoughts and that both are reasoning in line with accepted norms, the conclusions X reaches about what Y thinks (or will think) ought to be in good order.

Clearly, the products of this kind of thought replication process could be used in conjunction with the applications of the folk psychological framework in the course of making sense of intentional actions. But why isn't a pure co-cognition

account enough on its own for making sense of the action of others? The answer is quite simple. The folk psychological framework is necessary for understanding *reasons*, not just the interrelations between thoughts. And to understand a reason minimally entails an understanding of the interrelations between beliefs and desires (along with many other of their familiar partners).

A closing thought: It is surely an exaggeration to say that the loss of folk psychology would be 'the greatest intellectual catastrophe in the history of our species'. Still, for those who rely on it in order to make sense of intentional actions (our own and those of others), and who have based so many important practices on it, I don't doubt that its loss would be catastrophic. So: Long live Folk Psychology!

NOTES

[1] Terminology – or rather its associated effects – matters. Although many researchers from many different fields talk of 'folk psychology' and 'theory of mind' as if these labels had fixed designations, in fact they do not. Some only use them to denote a certain kind of practice or ability, while others refer to what putatively underlies and explains such practices and abilities. Thus it is often quite unclear exactly what one is committed to by saying that we get by in our everyday affairs by using folk psychology or a theory of mind. Loose talk sinks ships, so they say, but it can also keep them afloat. As long as our talk is unregulated it is impossible to assess claims properly.

[2] I have attempted to explain what lies at the root of folk psychological inabilities of sufferers with autism who fail to 'grasp the concept of belief' in Hutto (2006c).

[3] Vinden's cross-cultural studies, which employ four variants of location-change and false-belief tests, reveal significant variations in the understanding of belief between the children of certain cultures: "the response patterns vary from culture to culture, with the Western children the only ones who were at ceiling on all questions" (Vinden 1999, p. 32). In fact, in coping with the false-belief task – where children were asked what the protagonist would 'think' – those from the Mofu of Cameroon were only marginally above chance and those from the Tolai and Taine populations of Papua New Guinea were at chance. These results are even more remarkable given that the ages of the children involved included those of up to 8-years old (due to problems finding participants of the desired younger ages). Similar results of cross-culture comparisons concerning 'person' conceptions of a more general sort have "directly challenged the assumption of a single, universally applicable conception of the person and, perhaps, even more fundamentally, the view that treats the development of this conception as a straightforwardly individual and socioculturally decontextualized process" (Richner and Nicolopoulou 2001, p. 402).

[4] On its own this evidence does not provide a secure basis for an effective argument against the existence of theory of mind mechanisms (ToMMs). Drawing a comparison with culturally diverse folk theories of vision, Scholl and Leslie (1999, p. 137) insist that "Even specific beliefs about the concept of belief are not necessarily relevant: the concept of belief could be universally grounded in a module even though most cultures do not recognize the 'modular account' in their own folk psychology" (see also Mithen 2000b, p. 490). However, this cross-cultural divergence takes on new importance if independent arguments against the existence of ToMMs can be successfully mounted. It is a primary ambition of my recent book to supply such arguments (Hutto 2007a).

[5] As long as we do not equivocate in our use of terms, saying this does not imply that these others have a different folk psychology (stricto sensu). Claims about inter-cultural heterogeneity with respect to folk psychological practice must not be confused with claims that folk psychology is polymorphic, varying from culture to culture: it is not found in different *forms* elsewhere. The practice of making sense of intentional action in terms of reasons (*strictly defined*) simply could not be different in different parts of the world. This is, of course, wholly consistent with the possibility that the practice is not universal.

[6] In some cases the differences may be to do with frequency with which the folk psychological schema is used rather than the possibility of its use. For example, it has been shown that unlike North Americans,

the Chinese are less prone to explain events in terms of a person's reasons for acting or even by appeal to their personal character traits. Instead they prefer to make appeal to social-situational factors (Morris and Peng 1994). Thus when it comes to understanding what caused another to act the Chinese are more likely to cite such things as their 'being a victim of the Students' Educational Policy' or having 'recently been fired' rather than to cite features of the person's character or individual beliefs and desires (Lillard 1997, p. 271). Hence, it appears that favoured modes of action explanation, just like children's 'models of personhood', take "shape in an active interplay with culturally available models of personhood, which are not uniform either between or within societies" (Richner and Nicolopoulou 2001, p. 401).

[7] There are, of course, other non-folk psychological normalising explanations, but if the action was performed for reason only explanations that bring the folk psychological framework will do.

[8] It is worth saying something about the defining features of narratives *per se*. A very minimal definition will suffice. Lamarque tells us that for something to be a narrative "at least two events must be depicted in a narrative and there must be some more or less loose, albeit non-logical relation between the events. Crucially, there is a temporal dimension in narrative" (Lamarque 2004, p. 394, see also Lamarque and Olsen 1994, p. 225). This neutral characterisation easily lends itself to the idea that there are different types of narratives and that these can be classified by such common features as their constituents and subject matter. Folk psychological narratives – such as Little Red Riding Hood – are distinguished by being about agents who act for reasons. And, for my purposes, 'acting for reasons' is to be defined narrowly, as per tradition, as implying a belief/desire pairing consisting in or resulting in an intention.

[9] The practice of *providing* (or generating) narratives about reasons – the application of the derived framework – just *is* the practice of explicating and explaining action in terms of reasons. The success or otherwise of such explanations depends largely on who is doing the explaining.

[10] For example, within certain limits, propositional attitudes are subject to standard logical implications. If we know that 'X believes that P & Q', it can be safely assumed that 'X believes that P'. And propositional attitudes, or at least beliefs, are generative: It is possible to produce an indefinite number of new propositional attitudes by manipulating those one already has. This happens in a small way in every act of practical reasoning when new beliefs, desires or intentions are formed. If beliefs and desires stand in relation to sentences of some kind, all of this is easy to account for. Thus Fodor (1987, p. 71) is on the side of the angels in his long campaign for ensuring that we recognise the "independence of content from functional role".

[11] Or, for fans of dualistic approaches, that we can be related to sentences or that we can 'grasp' thoughts by means of them presents no interaction problem.

[12] Another way to achieve this would be to endorse the language of thought hypothesis. I argue elsewhere that we should not go down that road (Hutto 2007a). Supporters of TNL accounts typically hold that it is only after we achieve facility with the external symbolic forms of natural language – after we are practiced as engaging in 'public thinking' – that we eventually learn to 'think in our heads', using inner speech as a medium. This is an achievement not a given (Dennett 1998, p. 284). Private thinking requires replicating auditory and visual images of the structures used in overt speech acts and linguistic forms. Apparently school children do something similar when they first learn how to manipulate mathematical symbols publicly before being able to perform feats of 'mental mathematics'. Perhaps, the most remarkable example of this is the proficiency certain well trained Japanese children have for calculating enormously large sums using only a 'mental abacus'. The important point is that, on this account, it is either public natural language sentences or their internal proxies that serve as mediums for conducting propositional thinking.

[13] The idea of supermind takes its inspiration from Dennett's model of the conscious mind as having a stream-of-consciousness or Joycean character. On my rendering, like Frankish's, it is cast in a more dynamic role of a premising machine. Thus although its processing is still serial, it is used for engaging in deliberative acts of explicit practical or theoretical reasoning, those of the classical deductive, inductive and abductive variety. Thus Frankish (2004, p. 91) stresses it is "not just to speak as if the proposition were true but to reason as if it were – to take it as a premise". Thus supermental thinking involves "consciously and deliberatively calculating some of the consequences of one's premises. And acceptance – that is having a policy of premising – involves committing oneself to doing this, on appropriate occasions, in appropriate contexts" (Frankish 2004, p. 91). For this reason he recommends

that Dennett's proposed alias for conscious mind 'the Joycean machine' should be altered to that of 'the premising machine' in order to underscore its primary role. As this kind of reasoning activity is conducted consciously, it is easily accessible and familiar to us. This may also account for our mistaken tendency to see it everywhere, casting all cognition in its mould (Clark 1998, pp. 180–182).

[14] I diverge from Frankish's understanding of the basic mind because I deny that it should be understood in propositional attitude terms. Whereas he promotes a duplex approach – in which there are two types of propositional attitude at work – I retain his understanding of the supermental, but propose that it is only in the arena of ratio-discursive thinking that propositional attitudes are involved in cognition.

[15] A kind of instrumental thinking is possible without language, but augmenting our cognitive toolkit with a premising machine that makes use of linguaform structures would have constituted a major addition – one with truly transforming effects; not just a modest extension but a radical transformation of the cognitive possibilities. Nevertheless, the exercise of the recreative imagination would have prepared the grooves for and oiled the wheels of instrumental thinking using only non-sentential vehicles. This suggests a plausible explanation of how the wetware of the ancestral brain – which is not ready-made for logical reasoning but only imagistically grounded proto-logical thought – so easily accommodates supermental thinking. And this matters because the mere appearance of complex natural language forms is simply not sufficient to explain the ontogenesis of logical reasoning abilities. As Carruthers (1998, p. 108, emphasis original) observes, it is "quite obscure how the evolution of a grammar faculty could, by itself, confer capacities for non-demonstrative social, causal or explanatory reasoning". But if the basis of such reasoning was already familiar to certain non-verbal minds, it is easy to see how using public symbols with stable forms and content would have transformed it radically. Iconically based thinking has inherent limitations from which truly symbolic-based cognition does not suffer. Casting thoughts into a sentential format is precisely what is required for engaging in topic-neutral, domain-general reasoning. Even so, the capacity to perform supermental tasks does not come automatically; humans gain it only after having mastered the use of public language symbols. Clark is right to suppose that the acquisition of language allows for the 'ultimate upgrade' (Clark 1998, pp. 177, 179, 180). That said, it is important to stress that even in the human case, supermental thinking does not wholly usurp our more basic non-linguistic modes of 'reasoning' – it is not as if the acquisition of a premising machine effects a complete cognitive refit. Indeed, our older ways are likely to dominate in many circumstances with supermental capacities only being called on a fraction of the time. Basic minds, which come in different varieties, are good enough for getting most organisms through most situations, provided they are in their home environments. In the human case, they often allow us to navigate by autopilot as it were, but sometimes they are thwarted. Tackling problems that the basic mind cannot handle is a job for the supermind. However, contrary to what might be expected, going over to 'manual control' is to switch into low-gear thinking. It is a shift to a slow, careful deliberative mode that is serial, sequential and fragile. Clearly, the TNLH is therefore in line with dual-process theories of reasoning that have been advanced in order to make sense of the independent empirical data which shows that although sometimes people reason logically "sometimes they do not" (Gigerenzer 1997, p. 282; Over 2002, pp. 201–204). Speculatively, if reasoning by means of natural language structures did constitute such a vast improvement of our ancestors' cognitive possibilities this might potentially explain why modern humans are almost unique in the animal kingdom as having so completely out-competed all other species within its taxonomic family (Li and Hombert 2002, p. 176).

[16] One can endorse this account without taking sides in the debate about the 'location' of such cognitive processing. There is disagreement within the extended mind camp about whether or not supermental cognition involves an internal reconfiguration of procedures to which our biological brains are accustomed or merely the manipulation of external vehicles (Clark 1998; Dennett 1998; Mithen 2000a). What matters is that, on either account, by endorsing supermentalism one can be a 'realist' about beliefs and desires – accepting that there is a real difference between *really* acting for a reason and merely appearing to do so (e.g. by acting on the basis of intentional attitudes or even reflexively). Only those capable of bona fide practical reasoning using linguaform vehicles can act for reasons (non-verbal animals give the appearance of doing so, but this is an appearance: they are not 'true believers', still less 'true reasoners'). Hence, there is a way of agreeing with Davidson over Dennett without becoming Fodor: reasons *genuinely* figure in the causal explanation of intentional action (with some qualifications,

see Hutto 1999a). There are 'facts of the matter' about *whether* one has acted for a reason or not – and indeed about for *which* reason it was that one acted. It is an entirely different issue whether in any given case an interested onlooker or the person themselves might be in a position to say accurately *whether* they acted for a reason or for *which* reason they acted.

[17] It is easy to be misled on this score since the developmental psychology literature gives almost exclusive attention to the moment that children begin to pass false belief tasks. This can give the erroneous impression that folk psychological development culminates with an understanding of belief (Richner and Nicolopoulou 2001, p. 395). But folk psychological abilities do not spring into being complete as soon as children come by a basic understanding of belief. Indeed, the mere fact that we know that children reliably manage to pass false-belief tests at a certain age, under certain experimental conditions, does not tell us to what degree they understand that concept – nor anything about the true scope of their ability to apply it outside of such contexts. The myopia surrounding this particular experimental phenomenon has tended to blind researchers to the development of bona fide but nuanced folk psychological skills that only emerge after ages 4 and 5, as children hone their abilities. Thus "Proponents of the dominant theories have been notably quiet about what happens in development after the child's fifth birthday. However research that explores whether 5-year-olds can use simple false belief knowledge to make inferences about their own and other's perspectives finds that they singularly fail to do so" (Carpendale and Lewis 2004, p. 91).

[18] Culturally established texts of this sort are the most secure medium of introducing children to the folk psychological schema and training children in its application. Yet any story about reasons for action, even those related through casual conversations, has the potential to reinforce this understanding. And, of course, folk psychological narratives are most regularly relayed through conversation, despite the fact that the latter are less regimented and structured than the canonical texts used in much pre-school story telling. Like the well-constructed, familiar fairy tale cited above, conversations about reasons make mention of the labels of the attitudes and their appropriate object complements. Moreover, they serve to introduce these already familiar lexical terms and verbs in a new context. Nevertheless, everyday conversations describing reasons for action, unless they are well focused and extended, do not always reveal the *full* structure of reasons in the way more polished and detailed folk psychological narratives do. This is because our workaday folk psychological narratives are often truncated, in line with the rules of conversational implicature.

[19] A basic non-metarepresentational theory of mind, inherited along with rational theory construction mechanisms, is fundamental to this theory, which holds that our understanding of intentional actions done for reasons "appears to be constructed between 3 and 4" (Gopnik 1993, p. 332).

[20] The problem is well known to defenders of traditional, classical cognitivist approaches to artificial intelligence. It is widely accepted that if one requires explicit theoretical knowledge of the full range of possibilities that might obtain in a given domain in order to take appropriate action then – to sum up the worry with a slogan – agents would need to know practically *everything* in order to be able to do almost *anything*. As Varela et al. (1991) observe "after two decades of humblingly slow progress, it dawned on many workers in cognitive science that even the simplest cognitive action requires a seemingly infinite amount of knowledge, which we take for granted". More pithily, as Clark (1997, p. 6) remarks "a little reflection suggests that there would be no obvious end to the 'common-sense' knowledge we would have to write down to capture all that an adult human knows".

[21] Fodor (2000, p. 33) speaks of the 'ruinous holism' that follows from assuming "the units of thought are much bigger than in fact they could possibly be". In failing to face up to this challenge he accuses cognitive scientists of being in "deep denial" (Fodor 2000, p. 39). In an even gloomier assessment he writes "cognitive science hasn't even started; we are literally no farther advanced than in the darkest days of behaviourism" (Fodor 2000, p. 129). Carruthers is unimpressed by Fodor's arguments for the holistic character of domain general cognition. He claims they rest on a poor comparison with a rather specialised branch of cognition: scientific knowledge (Carruthers 2003, pp. 76–78). He also argues that the mind is much more modular than Fodor supposes.

[22] Heal (2000, p. 12) maintains that co-cognition involves seriously held beliefs entertained in hypothetical contexts, not 'pretend beliefs'.

[23] A consequence of co-cognition is that in understanding the thoughts of others we are afflicted with the 'Dr Watson constraint' (Botterill and Carruthers 1999, p. 90). At best, we are only ever able to *downgrade* our understanding of what others are likely to think, never upgrade it. Watson (not to mention LeStrade) always falls short of working out what Holmes thinks about any particular topic because of his limited deductive capacities and knowledge. It follows that if we understand the thoughts of others by using this method then we will always be constrained by the limits of our own intellectual capacities (and our ability to sensitively adjust these).

REFERENCES

Bartsch K, Wellman H (1995) Children talk about the mind. Oxford University Press, New York.

Bermúdez J (2003) The domain of folk psychology. In: O'Hear A (ed) Minds and persons. Cambridge University Press, Cambridge.

Botterill G, Carruthers P (1999) The philosophy of psychology. Cambridge University Press, Cambridge.

Bruner J (1990) Acts of meaning. Harvard University Press, Cambridge, MA.

Carpendale JLM, Lewis C (2004) Constructing an understanding of the mind: the development of children's social understanding within social interaction. Behav Brain Sci:79–151.

Carruthers P (1996a) Simulation and self-knowledge: a defence of theory theory. In: Carruthers P, Smith P (eds) Theories of theories of mind. Cambridge University Press, Cambridge, pp22–38.

Carruthers P (1996b) Language, thought and consciousness. Cambridge University Press, Cambridge.

Carruthers P (1998) Thinking in language? Evolution and a moduralist possibility. In: Carruthers P, Boucher J (eds). Language and thought: interdisciplinary themes. Cambridge University Press, Cambridge.

Carruthers P (2003) Moderately massive modularity. In: O'Hear A (ed) Minds and persons. Cambridge University Press, Cambridge.

Clark A (1989) Microcogniton. MIT Press, Cambridge, MA.

Clark A (1997) Being there. MIT Press, Cambridge, MA.

Clark A (1998) Magic words: how language augments human computation. In: Carruthers P, Boucher J (eds) Language and thought: interdisciplinary themes. Cambridge University Press, Cambridge.

Currie G, Ravenscroft I (2003). Recreative minds. Oxford University Press, Oxford.

Davies M (1998) Language, thought and the language of thought (Aunty's own argument revisited). In: Carruthers P, Boucher J (eds) Language and thought: interdisiplinary themes. Cambridge University Press, Cambridge.

Dennett DC (1998) Reflections on language and mind. In: Carruthers P, Boucher J (eds) Language and thought: interdisciplinary themes. Cambridge University Press, Cambridge.

Fodor JA (1983) The modularity of mind. MIT Press, Cambridge, MA.

Fodor JA (1987) Psychosemantics. MIT Press, Cambridge, MA.

Fodor JA (2000) The mind doesn't work that way: the scope and limits of computational psychology. MIT Press, Cambridge, MA.

Frankish K (2004) Mind and supermind. Cambridge University Press, Cambridge.

Gallagher S (2006) The narrative alternative to theory of mind. In: Menary, R (ed) Radical enactivism: focus on the philosophy of Daniel D Hutto. Consciousness and emotion John Benjamins: Philadelphia Amsterdam. pp 223–230.

Gallagher S, Hutto DD (2007) Primary interaction and narrative practice: from empathic resonance to folk psychology. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam.

Gigerenzer G (1997) The modularity of social intelligence. In: Byrne RW, Whiten A (eds) Machiavellian intelligence II: extensions and evaluations. Cambridge University Press, Cambridge.

Gopnik A (1993) How we know our own minds: the illusion of first-person knowledge of intentionality. In: Goldman A (ed) Readings in the philosophy of the cognitive sciences. MIT Press, Cambridge, MA (originally published in Brain Behav Sci 16.1–14, 1993).

Harris P (1996) Desires, beliefs and language. In: Carruthers P, Smith P (eds) Theories of theories of mind. Cambridge University Press, Cambridge, pp 200–220.
Heal J (1998a) Understanding other minds from the inside. In: O'Hear A (ed) Current issues in philosophy of mind. Cambridge University Press, New York.
Heal J (1998b) Co-cognition and off-line simulation: two ways of understanding the simulation approach. Mind Lang 13(4):477–498.
Heal J (2000) Other minds, rationality and analogy. Proc Aristotelian Soc, Suppl vol LXXIV:1–19.
Hutto DD (1999a) A cause for concern: reasons, causes and explanations. Philos Phenom Res 59(2):381–401.
Hutto DD (1999b) The presence of mind. Benjamins, Amsterdam.
Hutto DD (2004) The limits of spectatorial folk psychology. Mind Lang 19:548–573.
Hutto DD (2006a) Unprincipled engagements: emotional experience, expression and response. Target paper in Consciousness and emotion: special issue on radical enactivism, vol. 7.
Hutto DD (2006b) Embodied expectations and extended possibilities: reply to Goldie. In: Consciousness and emotion: special issue on radical enactivism, vol. 7.
Hutto DD (2006c) Four Herculean labours: reply to Hobson. In: Consciousness and emotion: special issue on radical enactivism, vol. 7.
Hutto DD (2007a) Folk psychological narratives: the socio-cultural basis of understanding reasons. MIT Press, Cambridge, MA.
Hutto DD (2007b) The narrative practice hypothesis. In: Hutto DD (ed) Narrative and understanding persons. Cambridge University Press, Cambridge.
Hutto DD (2007c) First communions: mimetic versus theory of mind abilities. In: Zlatev J, Racine T, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amerstdam.
Lamarque P (2004) On not expecting too much from narrative. Mind Lang 19:393–408.
Lamarque P, Olsen S (1994). Truth, fiction and literature. Oxford University Press, Oxford.
Lewis D (1970) How to define theoretical terms. J Philos 67:427–446.
Lewis D (1978) Psychophysical and theoretical identifications. Australas J Philos 50:249–258.
Li CN, Hombert J (2002) On the evolutionary origin of language. In: Stamenov M, Gallese V (eds) Mirror neurons and the evolution of brain and language. Benjamins, Amsterdam.
Lillard A (1997) Other folk's theories of mind and behaviour. Psychol Sci 8:268–274.
Lillard A (1998) Ethnopsychologies: cultural variations in theories of mind. Psychol Bull 123(1): 3–32.
Millikan RG (2004). Varieties of meaning: the 2002 Jean Nicod lectures. MIT Press, Cambridge, MA.
Mithen S (2000a) Paleoanthropological perspectives on the theory of mind. In: Baron-Cohen S, Tager-Flusberg H, Cohen D (eds) Understanding other minds. Oxford University Press, Oxford.
Mithen S (2000b) Mind, brain and material culture: an archeological perspective. In: Carruthers P, Chamberlain A (eds) Evolution and the modern mind: modularity, language and meta-cognition. Cambridge University Press, Cambridge.
Morris MW, Peng K (1994) Culture and cause: American and Chinese attributions for social and physical events. J Personality Social Psychol 67:949–971.
Nichols S, Stich S (2003). Mindreading: an integrated account of pretence, self-awareness and understanding of other minds. Oxford University Press, Oxford.
Over D (2002) Rationality of evolutionary psychology. In: Bermúdez J, Millar A (eds) Reason and nature: essays in the theory of rationality. Oxford University Press, Oxford.
Prinz J (2002). Furnishing the mind: concepts and their perceptual basis. MIT Press, Cambridge, MA.
Richner ES, Nicolopoulou A (2001) The narrative construction of differing conceptions of the person in the development of young children's social understanding. Early Educ Dev 12.
Scholl BJ, Leslie A (1999) Modularity, development and 'theory of mind'. Mind Lang 14:131–153.
Sellars W (1956/1997). Empiricism and the philosophy of mind. Harvard University Press, Cambridge, MA.
Varela FJ, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Vinden P (1996). Junin Quechua children's understanding of mind. Child Dev 67(4):1707–1716.
Vinden P (1999) Children's understanding of mind and emotion: a multi-culture study. Cogn Emotion 13(1):19–48.
Wellman H, Phillips A (2001) Developing intentional understandings. In: Malle B, Moses LJ, Baldwin DA (eds) Intentions and intentionality. MIT Press, Cambridge, MA.

VICTORIA McGEER

8. THE REGULATIVE DIMENSION OF FOLK PSYCHOLOGY

8.1. SETTING THE SCENE

A focal project in philosophy of mind and related cognitive disciplines is to understand the nature of our "folk-psychological" capacity for making sense of mind and behaviour – primarily human minds and behaviour, but also to some extent the minds and behaviour of other seemingly goal-directed organisms. As a species, we are, so far as we know, almost unique in our capacity to see individuals as 'minded' – as richly endowed with a panoply of intentional, emotional, perceptual and other kinds of states of mind that guide behaviour. Animal ethologists, particularly primatologists, continue to engage in a fascinating debate about whether and to what extent other species might share certain features of this capacity (see, for instance, Bekoff et al. 2002; Browne 2004; Byrne and Whiten 1988; Cheney and Seyfarth 1992; Heyes 1998; Pepperberg 1999; Povinelli 1996; Povinelli and Eddy 1996; Premack and Woodruff 1978). But one thing is certain, no other species has the sort of well-elaborated mentalizing skills that so importantly shapes our own experience in the world. We are not just mentalizers, we are inveterate mentalizers – finding it very hard to suppress our natural inclination to see various entities as humanly minded creatures, even when we have reason to worry or suspect that our well-elaborated practice of attributing psychological states is out of place on some occasions, except of course as a convenient façon de parler.

Of course, most of the time the suitability of such attributions doesn't worry us overmuch. We happily go about our daily lives commenting on the supposed mental lives of our infants, our family pets, the squirrels that chase one another about in the garden, even some of our own artefacts – computers, for instance, so far as they're running various interactive programs. All these things behave in a way – *act* in a way – that simply invites making sense of what they're doing in mentalistic terms – paradigmatically, theorists claim, in terms of beliefs and desires (the Ur mental states that together rationalize any apparently goal-directed behaviour). A nice demonstration of this mentalizing proclivity can be seen in typical subjects' reactions to Heider and Simmel's now famous stimulus: a short film involving three geometric figures that move around within and outside a partially enclosed rectangular space (Heider and Simmel 1944). The geometric figures, consisting of a circle, a small triangle and a larger triangle, look to be interacting with one another – with the large triangle "chasing" and then "bullying" the circle in and out of the rectangular enclosure, the smaller triangle "interceding to help the circle out", and then both circle and small triangle "making their escape" from the

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 137–156.

"persecuting attentions" of the large bullying triangle, which pounds against the rectangular enclosure in "anger and frustration", eventually destroying it. These folk-psychological attributions of intention, emotion, even virtues and vices within a sense-making narrative frame are incredibly robust across normal viewers, with individuals both spontaneously offering such narratives and also largely agreeing on how to interpret the behaviour of the "characters" involved. Moreover, at least in typically developing individuals, this tendency to be 'mentalistically responsive' to low-level perceptual cues seems to emerge very early on – well before, and so clearly foundational for, children's development of any elaborate, mentalistically characterizable social knowledge (see, for instance, Carpenter et al. 1998; Gergely et al. 1995; Johnson et al. 1998; Johnson 2000; Meltzoff 1995; Watson 1979).

Although tracing the earliest roots of our mentalizing proclivities is a fascinating endeavour (and one I will say a little bit more about at the end of section 3), my point here in emphasizing these very basic perceptual biases is to underscore how deeply we are biologically and socially primed to become, in the course of development, skilled and inveterate folk psychologists. And herein lies a problem. We become so thoroughly enmeshed in the *practice* of folk psychology – it becomes so thoroughly second nature to us – that pulling away from it enough to see its central features is surprisingly difficult to do. Over the years, philosophers, psychologists and other cognitive scientists have had furious and fruitful debates about many aspects of this practice, often drawing controversial and even counter-intuitive conclusions: for instance, that folk psychology is a commonsense proto-scientific theory of mind and behaviour applied to cases much like any other (proto)-scientific theory; that it involves a panoply of (theoretical) concepts (centrally belief and desire) that may well not survive in a mature science of mind and behaviour; that knowledge of our own minds is as theoretically mediated as our knowledge of other people's – hence, prone to the same sorts of errors demonstrably found in third-person attributions; that because it is acquired so early and with such little instruction, human beings must be innately pre-equipped with its basic concepts and/or causal principles, and so on. For each of these positions, there are strongly defended counter-positions, with a mixture of conceptual and empirical arguments offered on each side. Consensus on some issues is emerging slowly; on others, deep divisions remain. But amidst all the theoretical clamour surrounding folk psychology, there are central features of the practice that remain persistently invisible to all sides despite the fact that we live and breathe these features in our day to day lives.

The purpose of this paper is to render one of those invisible features visible to the theoretical eye. My aim is thus to present a somewhat heterodox image of folk psychology in contrast to what I will call the standard image – a term I will use to encompass significantly different views. The point of contention I want to mark is simply this. On the standard approach, the primary task for folk psychology is the *explanation and prediction of behaviour*. Some theorists will acknowledge that there is more to folk psychology than this – for instance, as folk psychologists we often evaluate individuals' agential capacities; we make judgements of rationality, of character and of responsibility; and we apportion various degrees of praise and

blame for their doings depending on the sorts of agential assessments we have made. But however central we folk may take these activities to be, they surely build on the capacity to attribute mental states in the first place; and we attribute mental states in the first place to try and figure out what others are up to – i.e. to try and explain and predict their behaviour. Theoretically, then, it seems entirely reasonable to focus on the capacity for explaining and predicting behaviour as the core capacity of interest. In any case, this assumption is not generally disputed. For the vast majority of theorists working in the philosophy of mind and other cognitive disciplines today, the aim has been to give an empirically well-supported and plausible account of our remarkable facility for everyday mentalistic explanation and prediction.[1]

The alternative image of folk psychology that I will present rejects the grounding assumption of this standard approach. But this bald statement must be carefully interpreted. It's not that I claim that there is no phenomenon of interest that cries out for theoretical explanation: we do have a remarkable facility for explaining and sometimes even predicting others' behaviour – particularly those we know well – by virtue of reasoning about their alleged mental states. I take this point to be uncontroversial. Nevertheless, I claim that by envisioning the goal of mental state attribution in such a narrow way, we overlook certain intersubjective features of the practice of folk psychology that make a critical difference for how we conceptualize the cognitive competence, or competencies, that underlie such folk-psychological skills. In particular, we overlook the way folk psychology operates as a *regulative* practice, moulding the way individuals act, think and operate so that they become well-behaved folk-psychological agents: agents that can be well-predicted and explained using both the concepts and the rationalizing narrative structures of folk psychology. Hence, on this alternative view, a central theoretical goal must be to give an empirically well-supported and plausible account of our remarkable ability to regulate others – as well as regulate ourselves – through the everyday practice of attributing mental states. Our capacity for "explaining and predicting" others' behaviour can then be understood more fruitfully in light of this account – viz. as a capacity that is exercised in more limited ways than envisioned by the standard approach, and as a capacity that has its ups and downs but which cannot be assessed, in any case, according to the standards of explanatory/predictive practices in the sciences.

But things are never quite as simple as they seem and before proceeding to the substance of the paper I should mention that there are two different versions in which the standard view of folk psychology comes. One is called the theory theory account, a term introduced by Adam Morton in 1980 to characterize a growingly popular view (Morton 1980); the other is the simulation account, originating in its contemporary form in the work of Alvin Goldman and Robert Gordon (Goldman 1989, 1992; Gordon 1986). According to the first view, folk psychology is primarily a theory of the antecedents of human behaviour, and the behaviour of intentional systems in general; it is a theory that depicts folk psychology as a theory designed in the ordinary scientific fashion to help us explain and predict one another (Churchland 1979; Dennett 1987; Fodor 1987; Lewis 1983). According to the second, simulation

view, folk psychology is primarily a method of explaining and predicting and is only incidentally associated with theorizing. The method consists in simulating the states of others, using ourselves as a model; it involves looking at the circumstances of others and then seeing what we would feel and think and want, and ultimately what we would do, in those circumstances (Currie and Ravenscroft 2003; Davies 1994; Goldman 1993; Gordon 1986; Heal 1998, 2003). There has been a long-running and enlightening debate between these two approaches and we shall be referring to them again in the course of the discussion. From the point of view of this paper, however, the commonality between the approaches is more important than the differences. For whether folk psychology is cast as a theory or as a method, it is seen in each case as primarily concerned with explanation and prediction. And that is the standard assumption that I contest.

As between the two versions of folk psychology, I should say that I think the theory theory comes across as rather better grounded. Folk psychology is distinguished by the fact that it provides us with a range of categories and kinds, including belief and desire, intention and action, deliberation and will, freedom and reason, emotion and valuation, and the like. And as between those different sorts of states and episodes, folk psychology gives us an overall framework according to which some of these are intelligible and predictable, others not. It is hard not to think of it, then, as a sort of theory, at least so long as we hold that our primary task in interacting with others is to explain and predict their behaviour. Simulation may be involved in the application of the theory, representing a heuristic for determining how the categories apply to another person. But it can hardly represent the whole of what folk psychology comes to. This being so, I shall concentrate in what follows on the theory theory version of the standard view.

The remainder of the paper is divided into two sections. In Section 8.2, I look at what I describe as the normative aspect of folk psychology, even under the standard view of it; this is particularly highlighted in the theory theory version. And then in Section 8.3 I go on to focus on how folk psychology proves not just to be a normative but a regulative practice.

8.2. THE NORMATIVE CORE OF FOLK PSYCHOLOGY

Even philosophers who support the standard view often support an assumption that will, in Section. 8.3, direct us towards the alternative, regulative conception of folk psychology. This section is given to examining this assumption, focusing on its centrality in the theory theory version and commenting on how it appears from the perspective of the rival, simulation story.

The assumption I have in mind is that when folk psychologists attribute beliefs and desires to predict and/or explain one another's behaviour, they are *making sense* of that behaviour in a distinctive way. They presuppose a model of what it is to be sensible, and they make sense of one another's behaviour so far as they succeed in squaring what is done with the requirements of the model. Folk

psychologists predict and explain *by* making attributions that make sense of one another's behaviour relative to a common fund of sense-making norms.

Even on the standard view, then, folk psychology is not just an explanatory/predictive practice; it is also, in a sense, a normative practice: a practice of showing how people's performances live up to certain norms and thereby become, in that special way, intelligible. Although folk psychologists may have some context-specific views about what others will do – based, for instance, on experience – the bulk of these views will be heavily influenced by norm-governed judgements about what others *ought* to do, what it makes sense to do, in the circumstances. They may know from experience with certain individuals that they are quirky and unusual in one or another respect and they will take account of such special knowledge in trying to give an account of the doings of people. But in general they will abstract from details of this kind and rely on the norms that people may generally be expected to satisfy.

Once the normative aspect of folk psychology is thus laid bare, two questions immediately arise: First, how do folk psychologists come to have sufficiently well-elaborated normative views about what others ought to think and do under various circumstances to drive their judgements in particular cases? And, second, how does it come about that these others generally think and do what they ought to think and do, so that making normative judgements about them works pretty well as a technique for explaining and predicting their behaviour?

A powerful and influential set of answers to these questions has been offered from within the theory theory camp of the standard view. These answers depend on the assumption that human beings are by and large *rational* creatures and, consequently, that the sense-making norms of folk psychology amount in good part to *norms of rationality*. Such norms will dictate what beliefs and desires should be formed in the presence of such and such bodies of evidence; how beliefs and desires should hang together in certain patterns of consistency and coherence; and what behaviour is suitable in the light of this or that set of beliefs and desires (Davidson 1984; Dennett 1987; Jackson 1992).

Given these assumptions we may now answer our two questions as follows: First, so far as developing well-elaborated views about what others ought to think and do under various circumstances, the epistemic task facing ordinary folk psychologists should not be as daunting as it might initially seem. They are, after all, rational creatures themselves, so will have an inbuilt sense of the kind of norms that govern rational thought and action. Specifically, folk psychologists will have an inbuilt sense of (1) the sort of beliefs and desires that rational creatures form under the circumstances in which they find themselves, and (2) the sorts of actions that rational creatures pursue in light of their beliefs and desires. And they will make their attributions of belief and desire, and their explanations and predictions of action, accordingly. With respect to the second question regarding how to account for the predictive/explanatory success of these normative judgements, the answer is equally clear. The targets of folk-psychological attention – viz. other agents – are

rational creatures too. Hence, they will generally think and do what they ought to think and do relative to the sense-making norms of folk psychology.

On this account, our capacity for folk-psychological explanation and prediction boils down to what Dennett has called a capacity to adopt the "intentional stance" (Dennett 1978, 1987).[2] The stance consists in viewing agents from within the assumption that they are rational, allowing for departures from rationality only so far as collateral evidence warrants; this will consist in evidence on human failures in general and evidence on the failures of those agents in particular. It means maximizing the extent to which they can be seen as exemplars of the rational model, within the constraints that such evidence imposes. But here the constraints are important too. As Dennett has repeatedly emphasized, since the thrust of folk psychology is to rationalize, the strategy only works so far as others' behaviour does not depart too wildly from norms of rationality.

Some have taken this feature as a reason to object to Dennett's way of accounting for our ordinary folk-psychological predictive-explanatory expertise. The concern has been that this approach is *overly* rationalistic, not allowing sufficiently for the myriad ways in which human beings lapse from the rational ideal – yet seem to remain perfectly understandable. Indeed, some simulation theorists have suggested that, since folk psychologists are perfectly able to take such foibles in their explanatory-predictive stride, this argues strongly in favour of their account of how folk psychologists arrive at sense-making judgements of others' behaviour – viz. by using themselves, cognitive warts and all, as a working model for exploring what others' will think and do (see, for instance, Gordon 1992).

There is a certain sense in which simulation theorists are right to worry about this aspect of Dennett's account, but they have focussed their worry in the wrong direction. And this is because they have been – like the theory theorists they criticize – thoroughly pre-occupied with the problem of explanation and prediction. It is, of course, true to say that human beings lapse from the ideal of rationality in all sorts of interesting ways – Dennett would be first to acknowledge this. But it would *not* be true to say, as simulation theorists have, that folk psychologists are very adept at predicting (or explaining) such lapses of rationality.[3] So they are no further ahead than Dennett in saying what is special or significant in the way folk psychologists deal with lapses of rationality – ways that allow them to take such lapses genuinely "in their stride", even if this is not a predictive-explanatory stride.

In my view, what is most noteworthy in these cases is the fact that folk psychologists have, as part of their overall competence, myriad techniques for identifying, excusing, blaming, accepting responsibility, apologizing and otherwise restoring confidence in the efficacy of the normative structures that govern the behaviour of individuals who *ought* to be explicable and predictable using the techniques of folk psychology, even though sometimes they are not. In other words, folk psychologists treat lapses of rationality, not just as "surd spots" in an explanatory/predictive theory, but as reasons to take some kind of remedial or restorative action. This suggests that the real problem with Dennett's "rationalistic" characterization of folk psychology's normative structure is not that it is overly demanding in terms of its rationality constraint, but rather

that it's overly austere in the assumption that the point of folk psychology, first and foremost, is to predict and explain. Thus, it gives a skeletal presentation of certain aspects of folk psychology's normative structure, but fails to take note of the regulative features that put a substantial kind of living flesh on the bare bones thus presented. I will return to this point in section 8.3 below.

In the remainder of this section I want to address a different question, related to the use of folk psychology, not just with one another, but also with other species – indeed, even with robots and computers. The discussion of this issue will help us to see the rival merits of the versions of the standard view but will also give us a nice segue into the discussion of the heterodox view in Section 8.3.

Recall the very striking phenomenon that I noted at the outset of this paper – viz. our readiness to see other entities as minded, so long as they display certain (low-level) agent-like characteristics. As folk psychologists we are primed to put our mentalizing skills to work in explaining and predicting the behaviour of all sorts of entities – and not just those that strike us as similar in form, features, habits or likely inclinations. This is surprising, since a priori one might have thought that little joy would come of this folk-psychological promiscuity. The selective pressures that gave rise to our folk-psychological capacities are likely to have been effective in the context of explaining, predicting and perhaps coordinating our behaviour with *conspecifics*; and so to have given rise to a cognitive mechanism evoked purely or primarily in our interactions with others of our kind. So why should we be prone to deploy our folk psychology with creatures of other kinds?

Simulation theory does not have a useful line on this question. On that theory, folk psychologists use their own cognitive mechanism to model others' cognitive situations, and so we should expect their predictive-explanatory expertise to be dramatically limited to creatures with relatively similar cognitive mechanisms. But while it does seem to be true that we have *more* explanatory-predictive purchase with conspecifics, we are not entirely baffled by creatures that are cognitively quite different from us. So the evidence suggests that no deep cognitive isomorphism is required to support such folk-psychological capacities.

Dennett's intentional stance approach offers a much better line on the question. On Dennett's view, folk psychologists can be expected to do a perfectly adequate job predicting the behaviour of a wide range of entities, both natural and artificial, because belief-desire psychology is abstract enough to track any pattern of behaviour that has a more or less rational profile (Dennett 1987, 1991). The precise nature of an entity's internal structure is neither here nor there so long as it manifests the requisite pattern of behaviour (cf. Jackson and Pettit 1990a). Hence, folk psychologists can use folk psychology to do predictive work, ascribing beliefs and desires in ignorance of the underlying cognitive (or mechanical) structures that are involved in supporting the predicted behaviour. Indeed, this capacity seems to give folk psychologists every reason to regard the entities so predicted as genuinely minded, as genuinely having beliefs and desires. For all it takes to have a mind – that is, to have beliefs and desires – from this perspective is to be "reliably and voluminously predicted from the intentional stance" (Dennett 1987).

Dennett's line on this issue is undoubtedly attractive but it runs foul of a singular fact that must now be put on the board. It's just this: Even though we may acknowledge that our predictive-explanatory talents extend easily to non-human entities, most of us remain unsure at the end of the day as to whether these entities are genuinely minded. Or at the least we are ambivalent on the issue. We may have some inclination to see our cats, our dogs and other sociable pets as genuinely minded. But we are much less committed to regarding the animals we eat, wild animals, and artificial entities (at least in our theoretical conception of them) as minded. Moreover, while the history of philosophy is replete with arguments that try to regiment our commonsense views about these matters, most of these arguments end by placing human beings firmly on one side of a deep metaphysical – or at least cognitive – divide, and animals and other non-human entities firmly on the other. We police the boundaries of "genuine" mindedness with a great deal of care and anxiety.

So now the question is why? What are we responding to, intuitively speaking, when we treat the sceptical problem of other (human) minds as a pleasing intellectual game, but the sceptical problem of other animal minds as a deep and difficult problem? Obviously, language must have a major role to play in the answer we give to this question. But it's important not to leap too quickly into thinking we have an answer just by adverting to language. For our linguistic capacities can be recruited to rather different explanatory ends. For instance, the fact that we can tell one another about our beliefs and desires will make it the case that we are much more readily predicted and explained from the intentional stance. So this may suggest that our intuitive resistance to treating non-human entities as genuinely minded boils down to the fact that other human beings are so much more 'reliably and voluminously' amenable to folk-psychological prediction than they are. But is this all that can be said? Or is there some special quality to be found in our folk-psychological interactions with one another because we are language users – in particular, *because we are adept in the use and understanding of folk-psychological concepts* – that is just not captured by this story?

I think there is and I turn to explain why in Section 8.3. But first I would like to make one last observation regarding simulation theory. One of the attractions of this approach is that it takes the special quality of our folk-psychological experiences with one another very seriously, aiming at an account that will do justice to it. But, once again, because simulation theorists are preoccupied with the issue of folk-psychological explanation and prediction, I think their concerns have taken them in exactly the wrong direction. Thus, they have tried to capture what is special about our folk-psychological interactions by identifying a *special sort of mechanism* for explanation and prediction that would give folk psychologists a peculiar intimacy with other human beings – viz. the mechanism of projective identification. But, as we saw above, this makes our predictive-explanatory success with creatures that are rather different from us rather difficult to explain.

For this reason we should take a different tack, locating the real problem with an intentional stance construal of our folk-psychological talents in quite a

different place. Specifically, the problem with this construal is not that it inappropriately *expands the scope* of our predictive-explanatory competence; it is rather that it inappropriately *collapses the range* of our folk-psychological talents into an explanatory-predictive competence, no matter how widely that competence might be seen to range. The role of explanation and prediction has been greatly over-exaggerated relative to its proper place in a more accurate account of the everyday interactions of ordinary folk psychologists.

8.3. FOLK PSYCHOLOGY AS A REGULATIVE PRACTICE[4]

To my mind, coming to see folk psychology as a regulative practice requires no great conceptual leap. Thus, even though I acknowledge that presenting it as such constitutes a heterodox theoretical view, my hope is to justify this theoretical re-orientation by showing how it follows straightforwardly from a few observations that many philosophers who work within the standard tradition should happily accept. As Wittgenstein would have said, the problem is simply that we have been transfixed by a theory, and the best way to recover from this condition is through an assemblage of reminders (Wittgenstein 1958).

Return for a moment to that pivotal moment in philosophy of mind when Wilfrid Sellars introduced his notorious myth of Jones, the genius amongst our Rylean ancestors. What was the point of proposing this myth? As Hutto (this volume) makes clear, it was to show how concepts like belief and desire that purportedly refer to hidden or 'private' psychological states in particular individuals could nevertheless be "primarily and essentially *intersubjective*" (p. 122). According to the myth, such terms are introduced, not ostensively, but functionally – as characterizing 'episodes' that purportedly play a role in individuals' psychological lives that is analogous to the role already played by various overt speech acts in their public lives of conversational exchange. As Sellars wrote:

> [In] the attempt to account for the fact that his fellow men behave intelligently not only when their conduct is threaded on a string of overt verbal episodes…but also when no detectable verbal output is present, Jones develops a theory according to which overt utterances are but the culmination of a process which begins with certain inner episodes…(His) model for these episodes which initiate the events which culminate in overt behaviour is that of overt verbal behaviour itself (Sellars 1997, p. 186).

Sellars' idea was to establish the intersubjective bona fides of intentional state terms ('belief', 'desire', 'thought', etc.) by drawing an analogy between the way these terms are introduced into the language and the way that scientific theory terms are – viz. by appearing in some plausible hypotheses about hidden features of entities that would account for their observable behaviour. However, in certain other ways, the analogy with scientific theory was not well chosen (cf. Hutto, this volume). In the first place, it casts Jones in the role of interacting with his companions as scientist to object, observer to observed, trying to understand what makes them tick. This suggests that the primary purpose for which Jones introduces these concepts – explanation and prediction – is the primary role they will play once they are put into intersubjective *use* – i.e. once Jones teaches his companions how to use this

'theory' in interpreting one another's behaviour. But consider how unlikely this is. After all – and this is the second point of disanalogy – the concepts Jones introduces are, as Sellars himself insists, concepts that have a particular kind of normative structure. For instance, as Sellars notes: "…in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state, we are placing it in the logical space of reasons, of justifying and being able to justify what one says" (Sellars 1997, p. 169). So clearly Jones and his companions will have a stake in the kind of characterizations they give of one another and in the kind of characterizations they make of themselves. Attributing a state of knowing is no light thing if suddenly it puts the person to whom it's attributed in the position of having to justify, or be able to justify, the various things that one says.

This suggests there ought to be a postscript to the myth of Jones. Maybe it goes something like this. Jones, having had his moment of genius, goes forth to teach his new theory to his contemporaries. They are quick studies, and soon come to understand what patterns of behaviour (overt and covert) are appropriately associated with various kinds of mental states. They get quite good at attributing beliefs and desires to one another for the purpose of explaining and predicting behaviour. And they even find they get quite good at making such attributions to themselves, giving their companions an enormous leg up in their predictive/explanatory endeavours. But after a while an odd thing begins to happen. They find that this new practice of psychological attribution is changing their lives in all sorts of subtle ways. Not only are they able to do a better job of predicting and explaining the behaviour of their companions, they find that others' behaviour and their own is starting to conform more neatly to patterns that are expectable under the attribution of various psychological states. For instance, they find that when they have publicly attributed a belief to themselves, they feel some pressure not to let their companions down in the expectations those companions now form about what they will say or do, and they find themselves responding to that pressure by monitoring what they say or do a little more carefully. Similarly, when others attribute a belief to them, they feel some pressure either to deny the belief, thereby refusing the normative weight of others' expectations, or to accept the attribution and live up to those expectations. In short, they have begun to experience the regulative power of the norms that surround these new folk-psychological concepts. And the upshot is that they no longer can interact with one another as scientist to object, as observer to observed, since the 'objects' themselves – viz. themselves and other agents – are changing under pressure of the "explanatory-predictive" attributions that are made to them.

Of course, this postscript to the myth of Jones is, like the original story, just a myth. As such, it is not meant to suggest anything about the natural evolution of folk psychology from a predictive-explanatory practice to that of a regulative practice. Rather, what I want it to highlight is the fundamental instability of conceiving of folk psychology right from the start as a proto-scientific practice akin to folk-physics, with the only difference between them being that one takes agents as the objects of predictive-explanatory concern and the other focuses on middle-sized, middle-distant physical objects. This analogy might work if all the objects

of folk-psychological attention were simply ignorant of our attempts to make sense of their behaviour by means of our rationalizing belief-desire attributions. But, unlike the non-human creatures to which we occasionally make such attributions, we human beings are not like that. We are neither ignorant of, nor indifferent to, the ways our thought and action get characterized in folk-psychological terms; and whether our aims are friendly or malicious, we have a stake in making ourselves comprehensible to others by way of shaping our own thought and action according to the sense-making norms of our shared folk-psychological practice. We also have a stake in encouraging others to make themselves likewise comprehensible.

But now what precisely is involved in making ourselves comprehensible according to the sense-making norms of a shared folk psychology? Is this equivalent to making ourselves into the image of rational agents, as the intentional stance view of folk psychology might suggest? Certainly it would seem that living up to norms of rationality is an important part of the self- and other-regulative skills we exercise in the context of our folk-psychological interactions. For instance, as I mentioned above, if we attribute a belief, either to ourselves or to others, then there are certain expectations that we form in light of that attribution – expectations of how we or they will behave; and these expectations are shaped by considerations of what it is rational to say or do in light of holding that belief.

Still, without discounting the rationality of many of our norm-governed interactions, this ideal seems far too austere to account for the myriad norm-governed expectations we develop around social behaviour, and the myriad norm-governed ways we learn to act so as to meet (and break) those expectations in sense-making ways. Is it *rational* to dress in a particular way when we appear before others in some authoritative role? In one sense, no. But it is a matter of social usage; so not dressing appropriately sends a message whether we intend it to or not. So it goes for countless other details relating to our daily interactions. Our ways of organizing our environment, our ways of conducting ourselves in spatial orientation to one another, our ways of using voice and body, our ways of dressing, all come to be normatively guided, conveying our thoughts and feelings to one another as much as our explicit communicative acts (Garfinkel 1967; Gergen 1982; Goffman 1959).

Skilled folk psychologists are aware of these nuances of minded behaviour and conduct themselves accordingly, observing or transgressing social norms and routines as suits their current purposes. On the one hand, many of our day-to-day transactions are made meaningful just by our conforming to such norms and routines. On the other hand, we often draw attention to ourselves by saying or doing things that are unexpected in context, creating 'surplus meanings' that others will respond to with interpretive efforts that are guided by an explicit use of the rationalizing (sense-making) apparatus of our shared folk psychology (Bruner 1990; Grice 1989). Thus, as skilled actors in the drama of normalized folk-interactions, we can also make deliberate *use* of unexpected sayings or doings to provoke others, not just to engage in rationalizing narratives that make sense of what we have said or done, but also that make sense of us in ways that we intend.

Of course, as with any complex skilled activity, degrees of proficiency in the normal and transgressive modes of folk-psychological interaction may vary widely, and in varying respects. For in fact, there are a number of skills that folk psychologists must draw upon in their everyday interactions with one another. First, there are skills involved in saying and doing what is generally regarded as normal, reasonable or expectable in context – knowing how to negotiate the complex norms that govern so many aspects of our social-communicative lives. And here the narrative structures of sense-making folk psychology have a role to play in establishing and reinforcing "canonical" patterns of behaviour: By way of these narratives, we learn what "reasonable" actors will think and do in a wide variety of circumstances (Bruner 1990). Still, reasonable actors are not limited to thinking and acting in canonical ways, as we have noted above. But there are skills involved in being transgressive as well – specifically, skills relating to the asking and giving of reasons for untoward behaviour that still manage to place such behaviour within the sense-making ambit of everyday folk psychology. Here the folk-psychological practice of attributing various psychological states finds a new role to play, not just in establishing what *is* canonical, but in negotiating what may count as reasonable even while departing from what is normally expected. The end result of such negotiations may be a reassertion of the canonical, with individuals either pleading special circumstances or conceding that they have not acted appropriately or reasonably; alternatively, there may be a more general revision of what folk psychologists should count as canonical under the circumstances in questions. Much will depend on the sort of folk-psychological explanations given and accepted by the principal actors in these negotiations (for a more nuanced account of these various folk-psychological skills, see Bruner 1990; Hutto 2004).

These observations are meant to highlight the fact, central to the regulative view, that skilled folk psychologists are not just able to read other people in accord with shared norms; they also *work* to make themselves readable in accord with those same norms. And, indeed, they are often inspired – or at least prodded – to do such work by the myriad ways, mentioned passingly in Sect. 8.2, that folk psychologists have to call one another to book when they have failed to perform as expected. This is one of the most telling features that differentiates folk psychology as a regulative practice from what it would be like if it were a mere explanatory-predictive practice, appropriately construed as a proto-scientific theory of behaviour. For in the case of a proto-scientific theory, failure in explanation and prediction should lead to some revision in the theory itself or in the way the theory is applied; it does not lead to putting normative pressure on the "objects" of theoretical attention themselves to encourage them to become more amenable to folk-psychological explanation and prediction on future occasions.

We thus have come to the core idea of the regulative conception of folk psychology. It is that our folk-psychological competence consists in our aptitude for making ourselves understandable *to* one another, as much as on our aptitude for understanding one another. And we do this by making (self and other) regulative use of the norms that govern appropriate attributions of a range of psychological states.

Thus, very often when we make such attributions to one another or to ourselves, we are not engaged in the activity of explaining and/or predicting behaviour at all. We are engaged in the activity of moulding behaviour – cajoling, encouraging, reprimanding, promising and otherwise giving ourselves over to the task of producing comprehensible patterns of well-behaved agency in ourselves and others from a folk-psychological point of view.

Still, it is one thing to lay out a heterodox theoretical position, and it is another to argue for its substantial merits. I do think there are many advantages to this theoretical reorientation, but in the interest of space, I conclude this paper by discussing only four of them. These four points build on each other according to the order in which I discuss them. The first two suggest a shift in the way theorists model individual folk-psychological capacities in order to account for the readiness with which we understand one another as mature folk psychologists; the third discusses how this kind of account connects naturally with an explanation of why folk-psychological knowledge of other human minds has a peculiarly intimate quality, more like projective attunement (as simulation theorists might say) than like theoretically mediated expertise; and, finally, the fourth point explores the developmental implications of this theoretical reorientation.

1. If we learn to govern our behaviour in ways that make us more readable to others, then their work as interpretive agents is greatly reduced. The same is true for us, if they learn to govern themselves likewise. This banal observation challenges an all too common assumption that understanding must require remarkable interpretive skills on the part of each individual if we are to explain the ease with which we ordinarily interact with one another. But just as one person's weight-lifting skills are not so remarkable if they lift a weight with others, so too a person's individual 'interpretive capacities' are not so remarkable if the burden of understanding is normally distributed between them and the person they come to understand (cf. Millikan 1993). We can, of course, show considerable interpretive ingenuity when called upon to do so; and this may require drawing upon fairly generalized knowledge about the psychological springs of human behaviour in addition to whatever particular knowledge we may have of individual peculiarities. However, what is exceptional about these moments is not just their relative infrequency, but also the difficulty and uncertainty with which such interpretive efforts proceed (cf. Hutto 2004). Moreover, if these moments become too frequent, we abandon our interpretive efforts altogether, adopting an 'objective' stance towards those who seem generally unresponsive to folk-psychological norms. We judge such individuals to be: 'eccentric', 'irrational', 'disordered', 'mad', 'compelled', 'discursively unreachable'. At the extreme, such individuals fall outside the realm of subjects we can interact with as free and responsible agents, able to make commitments to us or to understand the commitments we make to them (Bilgrami 1998; Dennett 1987; Pettit 1993; Strawson 1974).
2. If we make ourselves more readable to one another by conforming to shared norms of readability, it follows that much of the work of understanding one another in day-to-day interactions is not really done by us at all, explicitly or implicitly.

The work is done already and carried by the world, embedded in the norms and routines that structure such interactions (cf. Hutto, this volume). Hence, it is not just that we often behave in ways that make sense from a folk-psychological point of view; it is that many of our sense-making ways of behaving already have their significance built into them. Indeed, this foundation of pre-determined meaning dramatically expands our resources for what we can make meaningful, not just by ordinary recursive methods, but as already noted by creative transgression. That is, in breaking with norms and routines, we expect our actions to signal the need for special interpretation. But, equally, we generally only succeed in conveying what we mean when such interpretations can be reasonably guided by the meaning of whatever norms and routines are transgressed (metaphors, for instance, only work if the literal meaning of the words used serve as a plausible guide to what the speaker means). When we develop as folk psychologists, we no doubt hone our interpretive skills; but, more importantly, we come to live in a world where the kind of interpretive work we need to do is enormously enhanced by how much meaning our interactions already carry for us and carry because of the way we habitually conform to norms that invest our actions with common meaning. Becoming a skilled folk psychologist is, in this sense, no different from becoming a native speaker within a linguistic community. The ease with which we speak comprehensibly and understand others is based on the practices we share. Of course, the relationship between conforming to folk-psychological norms and conforming to linguistic norms is closer than mere analogy: in becoming proficient speakers of our native tongue, we become able folk psychologists, and vice versa. These two skills are importantly intertwined, since so many of our methods of being comprehensibly minded are embedded in the semantics and pragmatics of our language.

3. One of the complaints that simulation theorists have long made against rival theory theorists is that the latter make no attempt to account for the special character of our folk-psychological knowledge of other human beings. We seem to have a special understanding of the way they tick that is quite unlike our theoretical understanding of other objects, and even quite unlike our supposed folk-psychological knowledge of other non-human creatures. How is this special character to be explained?

   On the regulative view, I think the answer is straightforward. Folk-psychological expertise is *insider* expertise, the 'first-person' expertise of someone who is skilled at reading others in accord with shared norms because she is skilled at living herself in accord with those norms, and vice versa. As with the insider expertise of linguistic fluency, these two capacities come together. Indeed, it would be more accurate to say they are one and the same capacity exercised in two different respects: *speaking*, on the one hand, and *listening*, on the other – or, more generally, *expressing* and *attending to what is being expressed*. These are two sides of exercising a skill or competency; they are the two sides of what Gilbert Ryle called 'knowing how':

> If understanding does not consist in inferring, or guessing, the alleged inner-life precursors of overt actions, what is it? If it does not require mastery of psychological theory together with the ability to apply it, what knowledge does it require? We saw that a spectator who cannot play chess also cannot follow the play of others; a person who cannot read or speak Swedish cannot understand what is written or spoken in Swedish; and a person whose reasoning powers are weak is bad at following and retaining the arguments of others. Understanding is part of knowing *how*. The knowledge that is required for understanding intelligent performances of a specific kind is some degree of competence in performances of kind. The competent critic of prose-style, experimental technique, or embroidery, must at least know how to write, experiment or sew. Whether or not he has also learned some psychology matters about as much as whether he has learned any chemistry, neurology or economics. These studies may in certain circumstances assist his appreciation of what he is criticising; but the one necessary condition is that he has some mastery of the procedures, examples of which he is to appraise. For one person to see the jokes that another makes, the one thing he must have is a sense of humour and even that special brand of humour of which those jokes are exercises (Ryle 1949, p. 54).

Analysing normal folk-psychological competence as a kind of practical know-how makes its special character quite unmysterious. The way we 'get' what another person is up to is by knowing what it's like to be the kind of person whose sayings and doings are expressive of ways of being minded according to the norms we share. This attunement does not depend on putting ourselves in others' shoes. We are already in their shoes, as they are in ours. This doesn't mean that we can always express our folk-psychological know-how as others do. Hence their thoughts and actions may be surprising, intriguing, innovative, instructive from our own point of view. Nevertheless, they make sense to us because we have some competence in being a person *like that*: our understanding is schooled in precisely the way our own expressive performances are schooled, so we feel in our bones what it's like – what it *would* be like – to express ourselves in word or deed as they have done. Of course, others can sometimes act in ways that make no sense to us; but, then, so too can *we* sometimes act in ways that make no sense to us either. In both cases, our performances have failed to live up to norms that transform mere doings into actions that have meaning for us. In both cases our relation to the 'other' changes, from being someone on the inside familiar with the sense of agency expressed by our performances, to being on the outside where that sense of familiar agency fails. Needless to say, such failures are more disconcerting in our own case. But this is not because we have failed to *perceive* something that should be obvious to us from our first person point of view – viz. the causal springs of our own behaviour. It is because those ways of behaving, which we know to come from us, are not second nature to us *as ways of being minded*. Our ordinary competence for acting in comprehensibly self-regulated ways has somehow failed and we have limited resources for making sense of such failures except as departures from what we ought to do, and can work to try to do better in keeping with the normative dictates of our folk-psychological know-how (cf. McGeer 1996; McGeer and Pettit 2002).

4. Even supposing this skill-based account of folk-psychological expertise is on the right track, there remains the developmental question of how the norms which govern our shared ways of being minded become habitual for us, i.e. how they

become 'second nature'. Must we begin life with some innate sense of the special qualities of human behaviour in order to become conversant in the norms which govern our daily interactions? Or do we develop this sense as a consequence of becoming conversant in the norms? Here, too, a satisfying answer to such questions depends on keeping all parties involved in the process of normal psychological knowing clearly in view – namely, the child as developing folk psychologist and other people as the objects of her developing folk-psychological knowledge. For, as in the non-developmental context, there is work that must be done on each side in order for this kind of knowing to succeed, although the work that's done will naturally be of a somewhat different kind reflecting the peculiarities of the developmental situation.

To begin with the child as a developing folk psychologist, a number of empirical studies provide substantial evidence of an innate human disposition to respond differentially to social stimuli. From birth, infants will orient preferentially towards the human face and voice, seeming to know that such stimuli are particularly meaningful for them. Moreover, they register this connection actively, imitating a variety of facial gestures that are presented to them – tongue protrusions, lip pursings, mouth openings. They will even try to match gestures with which they have some difficulty, experimenting with their own faces until they succeed. When they do succeed, they show pleasure by a brightening of their eyes; when they fail, they show distress. In other words, they not only have an innate capacity for matching their own kinaesthetically experienced bodily movements with those of others that are visually perceived; they have an innate drive to do so. That is, they seem to have an innate drive to imitate others who they judge to be 'like me' (Meltzoff and Gopnik 1993; Meltzoff and Moore 1977, 1983, 1994, 1977).

Within a few months, infants will use this awareness of their essential link with others in yet more elaborate ways, imitating simple actions others perform on objects by nine months and more elaborate goal-directed activities by 18 months. Moreover, studies indicate that by 18 months babies are not just imitating what others actually do; they are performing their actions based on their understanding of what others mean to do. That is, they read through others' 'failures', improving on their actions in order to accomplish unmet, but apparently intended, goals (Meltzoff 1995). (For a more elaborate summary of this progression, see Gopnik et al. 2000.)

By this age, babies also show clear signs of using others' emotional responses to the world as a guide for their own behaviour, avoiding things that elicit fear, disgust or anger in others and approaching those in which others manifest interest or delight (Campos and Sternberg 1981; Repacholi 1998). They engage in 'joint attention' behaviours, following another's gaze or point to an object outside their visual field, and use pointing gestures themselves to direct another's attention in similar fashion. While some of these pointing gestures are 'instrumental', aimed at getting the object indicated, others seem clearly intended to do nothing more than elicit the other's response to something shared (Bates et al. 1975). In these ways and many others, even very young children show a basic readiness to learn from others' expressions and actions, interpreted therefore as having particular import

for themselves. As Bruner says, "we come initially equipped, if not with a "theory" of mind, then surely with a set of predispositions to construe the social world in a particular way and to act upon our construals. This amounts to saying we come into the world already equipped with a primitive form of folk psychology" (Bruner 1990, p. 73).

Now what about the objects of this primitive form of folk psychology? Though infants clearly respond differentially to social stimuli, it is crucial to keep in mind that they are helped along at every stage of this developmental trajectory by those who provide such stimuli. Human infants do not confront a world of 'unstructured experience', and not just because they have innate mechanisms for ordering whatever experience is given to them. Their own ordering capacities are given a significant boost, not just once but again and again over the course of development, by parents who shape their children's experience by involving them in structured interactions governed by the sense-making norms of folk psychology. That is to say, parents treat their children as intentional participants in practices that initially extend beyond their intentional competence, leaving the parents to maintain, and even exaggerate, the formal structure and affective import of such interactions for both. In fact, parents will often treat their children as initiating just such interactions, elaborating on what they do in ways that direct and enrich their children's initial intentions. Jerome Bruner has called this sense-making structuring of activity, 'parental scaffolding' (Bruner 1983). It begins in early infancy, when child and parent engage in 'conversational dances', trading vocalizations, gestures and expressions that the parent ensures are made 'conversationally relevant' to one another, not just by rhythm and affective tone, but often through responsive imitation (Brazelton and Tronick 1980; Kaye 1982; Trevarthen 1979). These mutual imitation games, delighted in by child and parent alike, are the primary means by which the child identifies him- or herself as like another and so, eventually, as a person whose thoughts and actions belong to the kind that persons produce (Meltzoff and Gopnik 1993). They are also the primary means by which the parent moulds the child to react, think and feel about things as persons do. As Meltzoff and Gopnik remark:

> …mutual imitation games are a unique and important constituent of early interpersonal growth. Adults are both selective and interpretive in the behaviour they reflect back to the child. They provide interpretive imitations to their infants, reflections that capture aspects of the infant's activity, but then go beyond it to read in intentions and goals to that behaviour…This, in turn, leads the infant beyond his or her initial starting point. Likewise, selected actions, especially those that are potentially meaningful in the culture, will be reflected back [to the infant] more often than others…(Meltzoff and Gopnik 1993, p. 349).

Thanks to these kinds of structured and progressively more sophisticated interactions with others, the experiences children have and the responses they are called to give shape their own sense of agency, both viscerally and conceptually. In the course of normal development, children are thus bootstrapped into regulating their own experiences, feelings, thoughts and actions, not just in concert with others, but in accord with the intersubjective norms of a shared psychological practice. In a word, they become comprehensible agents, i.e. good folk-psychological 'objects'; but the

manner in which they become such agents, no less than what they become, accounts in important ways for their capacity to understand others 'like them', i.e. others in whose image they have been substantially made.

NOTES

[1] For a few examples of theorists who depart from this norm, see Bruner (1990), Hutto (1999, 2004), McGeer (1996, 2001), McGeer and Pettit (2002), Pettit (1978).

[2] In what follows, I will focus on Dennett's account for simplicity's sake, but I take his views to be broadly representative of a family of functionalist views in the philosophy of mind.

[3] A closer examination of the social psychological evidence does not readily support this simulationist argument. In particular, there are many instances where failures of rationality do indeed subvert the folk psychologist's efforts to predict behaviour – and these failures are notable since folk psychologists seem to anticipate that others will behave more rationally than they actually do. In fact, folk psychologists seem to anticipate that they *themselves* will behave more rationally than they actually do once they are put into circumstances that had previously only been described to them (for instance, see Loewenstein and Adler 1995 on subjects' mis-anticipation of the 'endowment effect'; For a nice discussion of this point, see also Nichols 2003). The problem is: why should there be this bias towards rationality if folk psychologists are making their judgements by imaginatively projecting themselves (cognitive warts and all) into the relevant situations?

[4] Parts of this section draw on ideas I have developed in McGeer (2001, esp. pp. 117–123). For further reference, see also McGeer (1996) and McGeer and Pettit (2002).

REFERENCES

Bates E et al. (1975) The acquisition of preformatives prior to speech. Merrill-Palmer Q 21:205–26.

Bekoff M et al. (eds) (2002) The cognitive animal. MIT Press, Cambridge, MA.

Bilgrami A (1998) Self-knowledge and resentment. In: Wright C, Smith B, Macdonald C (eds) Knowing our own minds. Oxford University Press, Oxford.

Brazelton TB, Tronick E (1980) Preverbal communication between mothers and infants. In: Olson DR (ed) The social foundations of language and thought. Norton, New York.

Browne D (2004) Do dolphins know their own minds? Biol Philos 19(4):633–653.

Bruner J (1983) Child's talk: learning to use language. Norton, New York.

Bruner J (1990) Acts of meaning. Harvard University Press, Boston.

Byrne RW, Whiten A (1988). Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans. Oxford University Press, Oxford.

Campos JJ, Sternberg CR (1981) Perception appraisal and emotion: the onset of social referencing. In: Lamb ME, Sherrod LR (eds) Infant social cognition. Erlbaum, Hillsdale, NJ.

Carpenter M et al. (1998) Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. Infant Behav Dev 21(2):315–330.

Cheney DL, Seyfarth RM (1992) Characterizing the mind of another species. Behav Brain Sci 15:172–179.

Churchland P (1979) Scientific realism and the plasticity of mind. Cambridge University Press, Cambridge.

Currie G, Ravenscroft I (2003). Recreative minds. Oxford University Press, Oxford.

Davidson D (1984) Inquiries into truth & interpretation. Oxford University Press, Oxford.

Davies M (1994) The mental simulation debate. In: Peacocke C (ed) Objectivity, simulation and the unity of consciousness. Oxford University Press, Oxford.

Dennett D (1978) Brainstorms. Bradford Press, Montgomery, VT.

Dennett D (1987) The intentional stance. MIT Press, Cambridge, MA.

Dennett D (1991) Real patterns. J Philos:27–51.

Fodor J (1987) Psychosemantics. MIT Press, Cambridge, MA.
Garfinkel H (1967) Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ.
Gergely G et al. (1995) Taking the intentional stance at 12 months of age. Cognition 56(2):165–193.
Gergen KJ (1982) Toward transformation in social knowledge. Springer, New York.
Goffman E (1959) The presentation of self in everyday life. Anchor Books, New York.
Goldman A (1989) Interpretation psychologized. Mind Lang 4:161–185.
Goldman A (1992) In defense of the simulation theory. Mind Lang 7(1/2):104–119.
Goldman AI (1993) The psychology of folk-psychology. Behav Brain Sci 16:15–28.
Gopnik A et al. (2000) Early theories of mind: what the theory theory can tell us about autism. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ (eds) Understanding other minds: perspectives from developmental cognitive neuroscience. Oxford University Press, Oxford, pp 50–72.
Gordon R (1986) Folk psychology as simulation. Mind Lang 1:158–171.
Gordon R (1992) The simulation theory: objections and misconceptions. Mind Lang 17:11–34.
Grice HP (1989) Studies in the way of words. Harvard University Press, Cambridge, MA.
Heal J (1998) Co-cognition and off-line simulation: two ways off understanding the simulation approach. Mind Lang 13(4):477–498.
Heal J (2003) Mindreading: an integrated account of pretence, self-awareness, and understanding other minds. Mind 114(453):181–184.
Heider F, Simmel M (1944) An experimental study of apparent behavior. Am J Psychol 57:243–259.
Heyes C (1998) Theory of mind in nonhuman primates. Behav Brain Sci 21:101–148.
Hutto D (1999) The presence of mind. Benjamins, Amsterdam.
Hutto D (2004) The limits of spectatorial folk-psychology. Mind Lang 19(5):548–573.
Hutto D (2007) Folk-psychology without theory or simulation. In: Hutto D, Ratcliffe M (eds) Flok psychology re-assessed. pp 115–135.
Jackson F (1992) Block's challenge. In: Campbell K, Bacon J, Rhinehart L (eds) Ontology, causality, and mind: essays on the philosophy of David Armstrong. Cambridge University Press, Cambridge.
Jackson F, Pettit P (1990a) In defence of folk psychology. Philos Stud 57:7–30.
Johnson SC (2000) The recognition of mentalistic agents in infancy. Trends Cogn Sci 4(1):22–28.
Johnson S et al. (1998) Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. Dev Sci 1(2):233–238.
Kaye K (1982) The mental and social life of babies: how parents create persons. University of Chicago Press, Chicago.
Lewis D (1983) Philosophical papers, vol 1. Oxford University Press, Oxford.
Loewenstein G, Adler D (1995). A bias in the prediction of tastes. Econ J 105:929–937.
McGeer V (1996) Is "self-knowledge" an empirical problem? Renegotiating the space of philosophical explanation. J Philos 93:483–515.
McGeer V (2001) Psycho-practice, psycho-theory and the contrastive case of autism. J Conscious Stud 8(5–7):109–132.
McGeer V, Pettit P (2002). The self-regulating mind. Lang Commun 22(3):281–299.
Meltzoff AN (1995) Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Dev Psychol 31: 838–850.
Meltzoff AN, Gopnik A (1993) The role of imitation in understanding persons and developing a theory of mind. In: Baron-Cohen S, Tager-Flusberg H, Cohen DJ (eds) Understanding other minds: perspectives from autism. Oxford University Press, Oxford, pp 335–366.
Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. Science 198:75–78.
Meltzoff AN, Moore MK (1983) Newborn infants imitate adult facial gestures. Child Dev 54(3):702–709.
Meltzoff AN, Moore MK (1994) Imitation, memory and the representation of persons. Infant Behav Dev 17(1):83–99.
Meltzoff AN, Moore MK (1997) Explaining facial imitation: a theoretical model. Early Dev Parent 6:179–192.
Millikan RG (1993) White Queen psychology and other essays for Alice. MIT Press, Cambridge, MA.

Morton A (1980) Frames of mind: constraints on the commonsense conception of the mental. Clarendon Press, Oxford.
Nichols S (2003) Folk psychology. In: Nadel L (ed) Encyclopedia of cognitive science. Nature, London.
Pepperberg IM (1999) The Alex studies: cognitive and communicative abilities of grey parrots. Harvard University Press, Cambridge, MA.
Pettit P (1978) Rational man theory. In: Hookway C, Pettit P (eds) Action and interpretation: studies in the philosophy of the social sciences. Cambridge University Press, Cambridge, pp 43–63.
Pettit P (1993) The common mind: an essay on psychology, society and politics. Oxford University Press, New York.
Povinelli D (1996) Chimpanzee theory of mind? The long road to strong inference. In: Carruthers P, Smith P (eds) Theories of theories of mind. Cambridge University Press, Cambridge.
Povinelli D, Eddy TJ (1996) What young chimpanzees know about seeing. Monogr Soc Res Child Dev 61(3):1–190.
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 4:515–526.
Repacholi BM (1998) Infants' use of attentional cues to identify the referent of another person's emotional expression. Dev Psychol 34:1017–1025.
Ryle G (1949) The concept of mind. University of Chicago Press, Chicago.
Sellars W (1997) Empiricism and the philosophy of mind. Harvard University Press, Cambridge, MA.
Strawson P (1974) Freedom and resentment. Freedom and resentment and other essays. Methuen, London.
Trevarthen C (1979) Communication and cooperation in early infancy: a description of primary intersubjectivity. In: Bullowa M (ed) Before speech: the beginning of interpersonal communication. Cambridge University Press, New York.
Watson JS (1979) Perception of contingency as a determinant of social responsiveness. In: Thoman EG (ed) Origin's of the infant's social responsiveness. Erlbaum, Hillsdale, NJ, pp 33–64.
Wittgenstein L (1958) Philosophical investigations. Blackwell, Oxford.

JOSHUA KNOBE

9. FOLK PSYCHOLOGY: SCIENCE AND MORALS[1]

It is widely agreed that folk psychology plays an important role in people's moral judgments. For a simple example, take the process by which we determine whether or not an agent is morally blameworthy. Although the judgment here is ultimately a moral one, it seems that one needs to use a fair amount of folk psychology along the way. Thus, one might determine that an agent broke the vase *intentionally* and therefore conclude that she is *blameworthy* for breaking it. Here it seems that one starts out with a folk-psychological judgment (that the agent acted intentionally) and then uses it as input to a process that eventually yields a moral judgment (that the agent is blameworthy). Many other cases have a similar structure.

In recent years, however, a number of studies have shown that there are also cases in which the arrow of causation goes in the opposite direction. That is, there appear to be cases in which people start out with a *moral* judgment and then use it as input to a process that eventually yields a *folk-psychological* judgment (Knobe, 2003a, b, 2004, 2005a, b). These findings come as something of a surprise, and it can be difficult to know just what to make of them.

My own view is that the findings are best explained by the hypothesis that moral considerations truly do play a role in people's underlying folk-psychological concepts (Knobe 2003b, 2004, 2006). The key claim here is that the effects revealed in recent experiments are not the result of any kind of 'bias' or 'distortion.' Rather, moral considerations truly do figure in a fundamental way in the issues people are trying to resolve when they grapple with folk-psychological questions.

I must confess, however, that not all researchers in the field share this view. Although many have been convinced that moral considerations actually do play a role in folk-psychological concepts, others have suggested that there might be better ways to account for the results of recent experiments. What we are left with, then, is an increasingly complex debate. Critics of my original proposal have constructed alternative hypotheses that seem to account for all of the data without assigning any fundamental role to moral considerations. Defenders then conduct new experiments that appear to falsify these alternative hypotheses. But the critics inevitably respond by constructing even more sophisticated alternative hypotheses that manage to explain all of the new data while still assigning no fundamental role to moral considerations. And so the debate continues, with each new iteration yielding new theoretical insights and empirical discoveries.[2]

I will not be continuing that debate here. Instead, I want to focus on an issue that is somewhat broader and perhaps more basic. The critics sometimes seem to feel that moral considerations just *couldn't* be playing a fundamental role in folk psychology. The feeling is that, independent of the merits of any particular

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 157–173.

alternative explanation, one can tell that there must be *some* way to construct a valid alternative. This feeling is never articulated explicitly. Still, it comes through in the palpable sentiment that my defenders and I are upholding an absurd view and that we had really better come back to our senses.

My aim here is to confront that sentiment head on. In Section 9.1, I briefly review experimental evidence that suggests that people's moral judgments can sometimes affect their folk-psychological judgments. Then, in Section 9.2, I ask whether we have any general theoretical reasons to expect that moral considerations will not play any fundamental role in folk-psychological concepts.

9.1. I

Let us turn, then, to three folk-psychological concepts whose application has been studied experimentally. The first two have already been discussed in earlier papers and will only be described here in a highly condensed summary form. The third appears here for the first time, and I therefore discuss it in greater detail.

9.1.1. Intentional action

People ordinarily distinguish between behaviors that are performed *intentionally* (e.g., hammering in a nail) and those that are performed *unintentionally* (e.g., accidentally bringing the hammer down on one's own thumb). Clearly, this distinction sometimes has important implications for questions about moral praise and blame, but it is usually assumed that the distinction itself is a purely psychological one. Nonetheless, an ever-growing body of experimental evidence indicates that the moral status of a behavior can actually have an impact on whether or not people regard it as intentional.

The best way to demonstrate this influence of moral judgments on ascriptions of intentional action is to construct pairs of cases that are almost exactly alike but that differ in their moral status. Here is the first element in one such pair:

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.'

The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.'

They started the new program. Sure enough, the environment was harmed.

Faced with this first case, most people say that the chairman *intentionally* harmed the environment.

But now suppose that we create a morally good version by simply replacing the word 'harm' with 'help':

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.'

The chairman of the board answered, 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.'

They started the new program. Sure enough, the environment was helped.

When given this second case, most people say that the chairman *unintentionally* helped the environment. Yet the two cases are identical in almost all respects. It seems that the only major difference between them lies in the moral status of the agent's behavior.

In the years since this result was first reported, it has been replicated and extended in a wide variety of additional experiments. It has been shown that the effect continues to emerge when the stories are translated into Hindi and run on Hindi-speaking subjects (Knobe and Burra 2006), when the stories are simplified and given to subjects who are only 4-years-old (Leslie et al. 2006), and even when the stories are given to subjects who have deficits in emotional processing due to frontal lobe damage (Young et al. 2006).[3] At this point, no one doubts that people's use of the word 'intentionally' really is influenced by their moral judgments. The debate is simply about what this effect can tell us about the nature of folk psychology.

9.1.2. Reason explanations

Faced with this evidence that moral considerations play a role in people's application of the concept of intentional action, one possible response would be to deny that the concept of intentional action truly is a part of folk psychology. This response would allow us to hold on to the idea that morality plays no role in folk psychology, albeit at the expense of forcing us to admit that our intuitive notion of the scope of folk psychology was not quite correct. To me at least, this response seems a bit desperate, and no one has actually argued for it in print. Still, it comes up often in conversation, and as experimental research continues to show new ways in which the concept of intentional action is sensitive to moral considerations, it may come to seem more and more plausible.

It can be shown, however, that similar effects arise even for concepts that are undeniably folk-psychological. Thus, consider the practice of explaining behavior using *reasons*. A clear example would be the sentence:

He went to the kitchen to get a beer.

This sentence explains an agent's behavior ('He went to the kitchen…') by giving his reason for performing it ('…to get a beer'). Here we seem to have a prototypical case of a folk-psychological judgment. No one would claim that explanations of this type belong to the domain of moral cognition.

And yet, it can be shown that moral judgments actually affect people's use of reason explanations (Knobe 2004). Indeed, the effect can be seen in the very same pair of vignettes we used above. Faced with the first vignette, most people think it sounds right to say:

The chairman harmed the environment in order to increase profits.

But faced with the second vignette, most people *don't* think it sounds right to say:

The chairman helped the environment in order to increase profits.

This pattern of results suggests that people's use of reason explanations is actually sensitive to moral considerations.

It is not known precisely why this effect arises. One plausible hypothesis would be that people are using the concept of intentional action in the process by means of which they evaluate reason explanations. Perhaps people only accept reason explanations for behaviors that they regard as intentional. Then, since moral considerations play a role in people's concept of intentional action, they end up playing a role (somewhat indirectly) in the practice of reason explanation.

9.1.3. Valuing

People ordinarily distinguish between *desiring* and *valuing*. Thus, when a heroin addict is roaming the streets looking for his next fix, we might say that he 'wants' the fix but not that he 'values' it. And we would say the same about the man on a diet who feels overwhelmed by an urge to have another slice of chocolate cake. Philosophers typically find that they all share the same intuitions about how to apply the concept of valuing in cases like these, but it has proved notoriously difficult to say anything very definite about the basic criteria underlying these intuitions. One wants to know exactly how people go about distinguishing values from attitudes of other kinds.

This question has not received much attention from researchers in folk psychology, but it has been discussed extensively within a certain tradition in moral philosophy. This tradition begins with Watson's influential claim that

an agent's values consist in those principles and ends which he – in a cool and non-self-deceptive moment – articulates as definitive of the good, fulfilling and defensible life (Watson 1975: 215).

Watson later retracted that claim, worrying that it conflated the notion of valuing with the notion of judging something to be good (Watson 1987). But in the years that followed, a number of other philosophers have offered competing accounts.[4] We will not be concerned here with the differences among these various proposals. Instead, the focus will be on the assumption, shared by all of the views proposed thus far, that the concept of valuing can be defined in purely descriptive, non-normative terms.

I had never thought to question this assumption until the philosopher Erica Roedder suggested to me (in conversation) that there might be more to the story. She pointed out that the ordinary distinction between desiring and valuing might be bound up in a fundamental way with certain *moral* questions. So, for example, when we are trying to determine whether or not the heroin user 'values' his next fix, it might be that we are not simply concerned with purely descriptive questions about the nature and functional role of the user's attitude. Perhaps our reluctance to classify this attitude as a 'value' is due in part to our sense that heroin truly *is* a bad thing.

One way to make sense of this hypothesis is to suppose that the concept of valuing is a prototype concept. In other words, we can suppose that the concept of valuing is represented by a cluster of features, such that no individual feature is strictly necessary but each feature has been assigned a certain weight. If a particular

attitude shows enough of the relevant features, it will be classified as one of the agent's 'values.' It would be extremely difficult to provide an exhaustive list of the features that play a role here, but we can easily list a few that are likely to be relevant. When people are trying to determine whether or not the agent values a certain object *o*, they probably consider psychological features like:

- whether the agent has a conscious belief that *o* is good
- whether the agent is motivated to promote *o*
- whether the agent experiences guilt when she fails to promote *o* in circumstances where she could have
- whether the agent has a second-order desire for *o* (i.e., a desire to desire *o*).

Each of these psychological features has a certain weight. But the psychological features are not the only features of the concept. There is also a moral feature, namely, *whether the object o truly is morally good.*

Now, clearly, it would be foolish to suggest that moral goodness is a necessary condition in our concept of valuing. But that is not the claim under discussion here. The claim is simply that moral goodness has a certain *weight* in the process of classification. If an agent has all of the relevant psychological features, this extra weight simply won't be needed. The psychological features prove sufficient all by themselves. So the only way to see the significance of the moral feature is to look at cases where the agent has some of the psychological features but lacks others. In cases like these, the psychological features will not be sufficient all by themselves. The attitude needs the moral feature before it has enough weight to push our intuitions over the critical threshold.

Together, Roedder and I conducted an experiment to test this hypothesis. All subjects were given a story about an agent who has some of the relevant psychological features but lacks others. (In our story, the agent has motivation and guilt but not conscious belief or second-order desire.) The key question was whether people's classification of the agent's attitude would be influenced in any way by the perceived moral status of its object.

Subjects in one condition were given a story in which the agent feels a certain pull toward actions that would normally be perceived as *morally good*:

> George lives in a culture in which most people are extremely racist. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of people of his own race at the expense of people of other races.
>
> Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he harms people of other races. And sometimes he ends up acting on these feelings and doing things that end up fostering racial equality.
>
> George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial equality and just act to advance the interests of his own race.

After reading this story, subjects were asked whether or not they agreed with the sentence: 'Despite his conscious beliefs, George actually values racial equality.'

Subjects in the other condition were given a story that was very similar to the first one but in which the agent feels a pull towards actions that would normally be perceived as *morally bad*:

George lives in a culture in which most people believe in racial equality. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of all people equally, regardless of their race.

Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he helps people of other races at the expense of his own. And sometimes he ends up acting on these feelings and doing things that end up fostering racial discrimination.

George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial discrimination and just act to advance the interests of all people equally, regardless of their race.

These subjects were then asked whether or not they agreed with the sentence: 'Despite his conscious beliefs, George actually values racial discrimination.'

This experiment provides an initial test of our hypothesis. The attitudes depicted in the two stories differ in their moral significance, but they seem not to differ in any of the relevant psychological features. In both cases, the agent has motivation and guilt but not conscious belief or second-order desire. Yet, despite this similarity in psychological features, we find a marked asymmetry in people's intuitions. Subjects were significantly more inclined to say that the attitude was one of the agent's values in the morally good case than they were in the morally bad case. This result provides some tentative support for the view that moral judgments actually do play a role in people's concept of valuing.

9.1.4. Summing up

The results described here appear to indicate that people's applications of folk-psychological concepts can sometimes be influenced by their moral judgments. These results therefore provide some initial support for the claim that moral judgments are actually playing a role in people's folk-psychological concepts themselves.

But, of course, one cannot infer directly from the conditions under which a concept is applied to the structure of the concept itself. It is always possible that we will be able to come up with an alternative explanation that accommodates all of the relevant data without according any fundamental role to moral considerations in our underlying folk-psychological concepts. Perhaps the results described above are simply due to conversational pragmatics, emotional biases, or some other factor that has nothing to do with the underlying structure of people's concepts. A number of researchers are actively pursuing explanations along precisely these lines (see, e.g., Adams and Steadman forthcoming; Malle 2006; Nadelhoffer 2004, 2006; Nichols and Ulatowski 2006), and nothing I have said here provides any evidence against their hypotheses. Ultimately, the only way to assess these alternative explanations is to engage in a detailed examination of the existing experimental data.[5]

I will not be taking up that task here. Instead, I will be concerned with the initial motivation that leads researchers to search for alternative explanations in the first place. There seems to be a widespread intuition that moral considerations just *couldn't* be playing any fundamental role in people's folk-psychological concepts and that it therefore *must* be possible to find some other way of explaining the data. This intuition does not appear to depend on the evidence for any particular alternative hypothesis. It seems to stem instead from a more general theoretical commitment.

Clearly, the commitment here is not to the idea that moral considerations never play a fundamental role in any of our concepts. It is usually assumed that moral considerations do play a role in the concepts of blameworthiness, fairness, etc. and the researchers pursuing alternative explanations for the data described here do not seem to feel compelled to search for alternative explanations in those other cases as well. So the thought seems to be that there is something special about folk-psychological concepts in particular which makes it implausible that moral considerations could play any fundamental role in them. What I want to ask now is whether there really are any general theoretical reasons for holding this view.

9.2. II

Much of the attractiveness of the view appears to stem from the idea that folk psychology is in some important way similar to *science*. This idea is never spelled out explicitly, but the underlying argument seems to run something like this:

(1) Folk psychology is similar in many ways to a scientific theory.
(2) Scientific theories do not classify objects based on their moral properties.

We therefore have good reason to suppose that:

(3) Folk psychology does not classify objects based on their moral properties.

Of course, this is a not deductively valid argument, but it is a powerful one all the same. Both of the premises seem initially plausible, and together they appear to provide strong evidence for the conclusion.

To get a sense for the basic idea behind premise (2), it may be helpful to consider an example. Suppose we were able to observe a team of physicists studying the trajectories of certain projectiles. We might expect them to classify a projectile in terms of its mass, velocity, direction, and so forth. But suppose we then discover that their judgments can actually be influenced in some subtle way by *moral* properties, so that they sometimes end up applying scientific concepts to a projectile differently depending on whether they believe that it was morally right or morally wrong to launch it in the first place. In such a case, we surely would not conclude that moral properties actually play some important role in the basic concepts of physics. Instead, we would assume that the physicists were subject to some kind of bias that distorted their scientific judgment.

In thinking about cases like these, we brush up against some difficult questions about the relationship between science and morals. Someone might argue that initial impressions are deceiving here and that there really is some subtle sense in which

scientific theories end up classifying objects on the basis of their moral properties. Perhaps there actually is something to this charge, but let us put it to the side for the moment. For the sake of argument, we can simply assume that scientific theories do not classify objects on the basis of their moral properties. Then we can go on to ask what implications this putative fact about scientific theories might have for the study of folk psychology.

The key move, then, is from the claim that moral considerations are excluded from certain aspects of scientific theorizing to the claim that moral considerations are excluded from parallel aspects of folk psychology. This move rests on a certain analogy between science and folk psychology. The view is that, although science is more rigorous, more systematic and more explicit, we have reason to expect that the most basic practices associated with science will be found in folk psychology as well.[6]

It is this view that I want to examine here. To address these issues, we need to look more closely at the role science plays in people's lives and the factors that have made it such a dominant approach to systematic inquiry. Then we can check to see whether those same factors can be found in the case of folk psychology or whether folk psychology differs from science in some important respect.

1. Contemporary enthusiasm for the analogy between folk psychology and science appears to stem, at least in part, from the extremely salient position that science occupies in modern life. Everywhere one looks, one finds the fruits of scientific inquiry, and it is easy to find oneself thinking that the practices we now associate with science are in some way 'natural' to human beings. One almost finds it difficult to imagine any other way of generating predictions or explanations.

But, of course, the matter is not so simple. Many of the practices that we now associate with science arose in a particular cultural context in the not-too-distant past. These practices are now quite widespread, but one cannot therefore infer that they reflect anything fundamental about human nature. It may well be that they only came to occupy such a salient position in our society because they do such a good job of solving the kinds of problems we most often encounter in modern life.

Perhaps some of the confusion here arises from our tendency to lump together a diverse array of practices and label them all collectively as 'science.' Some of the practices that fall under this label really do seem to reflect something fundamental about human nature. These practices can be found in young children and in people from other cultures, and many cognitive scientists believe that they have an innate basis (see, e.g., Bloom 2004; Gopnik et al. 2004; Keil 1989; Pinker 1997). But not all of the practices associated with science work like that. Some of them were only developed in recent centuries and appear to be passed down from one generation to the next through explicit instruction. There is little reason to suppose that these practices reflect anything fundamental about our innate cognitive endowments (Faucher et al. 2002; McCauley 2000).

The thing to keep in mind in discussing practices of this latter type is that they arose as a result of certain contingent historical events. There is an important sense in which the 'scientific revolution' of the sixteenth and seventeenth centuries truly was a *revolution*. It introduced genuinely new practices, practices that cannot be

found in earlier eras. These practices subsequently assumed a dominant role in the kinds of inquiry conducted in systematic research programs, and we have ample evidence that they do a wonderful job of helping us get at the truth about certain difficult questions. But it would be wrong to suppose that there is something basic about human nature that compels us to adopt these practices in the form in which they presently exist. At other times and in other cultures, people have generated predictions using approaches that differed in various ways from the approach we now associate with science.

With this background in place, we can return to our central question. That question was whether we have any general theoretical reason to suppose that folk psychology treats moral considerations in the same way that science does.

2. The idea that folk psychology might be similar to science has been encouraged by the claim that folk psychology should be understood (in a certain technical sense) as a *theory*. The association here is understandable. As soon as one hears the word 'theory,' one immediately thinks of the sciences. So when one is told that folk psychology itself should be understood as a theory, one naturally leaps to the conclusion that folk psychology should be understood as something like science. It is therefore essential to remember that the word 'theory' was first introduced into this discussion in a highly specialized sense that did not carry any implications about all of the practices we normally associate with science.

The idea that folk psychology should be understood as a theory was first developed by Sellars (1956) and then entered the world of cognitive science through the influential work of Premack and Woodruff (1978). These researchers were concerned with the fact that folk psychology doesn't just give us a collection of empirical generalizations about observable phenomena but actually provides a deeper sort of account that works by explaining observable behaviors in terms of unobservable mental states. As Premack and Woodruff put it:

> In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and to others…A system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms (Premack and Woodruff 1978, p. 515).

I have no objections to this use of the term 'theory,' but when the term is used in this way, one cannot simply assume that every theory is best understood on the model of science. After all, a system of belief can easily qualify as a 'theory' in Premack and Woodruff's sense even if it does not have many of the properties we normally associate with scientific inquiry. To take a particularly glaring example, certain *religions* posit unobservable entities that can be used to predict observable events and might therefore be described as 'theories.' Now, it does seem fair to say that a religion can offer us a theory about how the world works, but one sees immediately that the theories offered by religions differ from scientific theories in a number of important respects.

In particular, the argument sketched above seems to depend in a crucial way on the distinctive features of *scientific* theories. There is some intuitive plausibility to the inference: 'Folk psychology is similar to science. Therefore, it does not

classify objects based on their moral properties.' But the argument loses all its force when we change it to: 'Folk psychology is similar to religion. Therefore, it does not classify objects based on their moral properties.' Religions serve a great many different functions in our lives, and prediction is just one. No one would be surprised to find that religious theories are connected in an essential way with moral considerations.

In short, it is easy to get confused by the claim that folk psychology is a 'scientific theory.' We really need to divide this claim into two parts – the claim that folk psychology is a *theory* and the claim that folk psychology is *scientific*. The claim that folk psychology is a theory simply isn't very relevant to the questions we are trying to address here. What we really want to know is whether folk psychology is, in the relevant sense, scientific.

3. Our concern, then, is with the distinctive features of scientific theories – the features that distinguish scientific theories from theories of other types. It seems that these features lie not so much at the level of content as at the level of methodology. The methods we use to evaluate scientific theories seem to differ in some important respects from the methods we use to evaluate theories of other types.

Perhaps the most striking aspect of scientific methodology is its sensitivity to empirical evidence. We use scientific theories to generate predictions, which can then be tested through observation or experiment. Theories that yield false predictions may be revised or abandoned. So one way to determine whether folk psychology is something like a scientific theory would be to ask whether it, too, is sensitive in the right way to empirical evidence.

A whole industry of research has arisen to answer this question, and a wide variety of competing theoretical frameworks have now been proposed. Some have argued that people can revise the basic framework of folk psychology using the very same psychological processes that scientists use to revise their theories (e.g., Gopnik and Wellman 1992); others argue that the basic framework underlying folk psychology is innate and is only sensitive to empirical considerations through a process of evolution by natural selection (e.g., Baron-Cohen 1995); and still others have suggested that folk psychology might be subserved by an innate module that uses empirical evidence to set certain highly specific parameters (Stich and Nichols 1998). The debate among these various positions is still ongoing.

But I worry that this research does not really get at the question we are trying to address here. It is not as though scientific theories are the only systems of thought that prove sensitive to empirical considerations. One finds at least some level of sensitivity to empirical considerations even in systems of thought that are clearly non-scientific. Consider a simple example. In the seventeenth century, many European Jews believed that Shabbatai Zvi was the messiah. They then received a shocking piece of disconfirming empirical evidence (Shabbatai Zvi converted to Islam), and most of them soon abandoned their previous belief. What we have here is a clear case of a group of people revising their views in light of empirical evidence. But no one would suggest that the followers of Shabbatai Zvi were propounding a genuine scientific hypothesis! Clearly, their belief was a religious

doctrine, and the criteria used to evaluate it therefore differed quite radically from the criteria we typically find in scientific inquiries.

The key mistake here is to assume that we can figure out what is special about scientific inquiry simply by looking at the considerations that scientists normally take into account. This approach has undoubtedly yielded many important insights, but it is not sufficient all by itself. We also need to look for kinds of considerations that scientists *don't* take into account. That is, we need to look for kinds of considerations that figure prominently in other systems of thought but do not play any role in scientific inquiries.

To get a sense for what I mean here, consider the many kinds of criteria we might use in deciding between competing religious doctrines. It seems that many of these criteria play no role at all in scientific investigations. Indeed, one of the key turning points in the scientific revolution was the struggle to establish a special realm of inquiry from which these other criteria would be completely excluded.

For present purposes, one of the most important distinctions between scientific and non-scientific theories lies in the differing roles they assign to *moral* considerations. We expect a religious doctrine to give us some measure of moral guidance, and if it fails to do so, we regard it as deficient in an important respect. By contrast, when we are evaluating a scientific theory, it seems that we are *not* supposed to be concerned in an essential way with moral questions. The theory can be perfectly successful from a scientific point of view even if it provides no moral guidance at all. In fact, we might find that the theory carves up the phenomena in a way that is completely orthogonal to the categories that prove most relevant in our moral thinking. But no matter. As long as the theory does well according to the distinctive criteria of science (empirical adequacy, simplicity, etc.), we are supposed to consider it a success.

We can now get a better handle on the question as to whether or not folk psychology is something like a scientific theory. In addressing this question, it is not enough just to ask whether or not folk psychology is sensitive in the right way to the kinds of considerations that play a role in scientific inquiry. We also need to know whether it resembles science in *excluding* the kinds of considerations that are usually excluded from scientific inquiries.

4. At this point, it might be thought that we really do have quite good reason to assume that folk psychology excludes the very same sorts of considerations that are normally excluded from scientific inquiries. After all, it is a conspicuous fact about our modern age that scientific approaches have proved extraordinarily successful in the systematic research programs where they are most commonly employed. One might therefore be tempted to conclude that the most effective way to proceed as folk psychologists would be to use almost exactly the same methods we find used in scientific inquiry.

But perhaps this conclusion is a bit too hasty. Clearly, there are some important differences between what we are looking for in a scientific research program and what we are looking for in a folk theory like folk psychology. So it is at least conceivable that the approach that best serves our needs in scientific research

programs will not also best serve our needs in folk theories. Before we can determine whether or not there is reason to suspect that folk psychology uses a scientific approach, we therefore need to look in more detail at the advantages and disadvantages of that approach more generally.

One of the chief advantages of the scientific approach is its unparalleled predictive power. By excluding many of the criteria used in other kinds of inquiry, a scientific investigation can arrive at theories that do an extraordinarily good job at predicting the phenomena under study.

But this predictive power comes with a price. A scientific theory is a highly special-purpose tool. It might do an excellent job when our aim is to make predictions, but it won't necessarily prove helpful in all of the other tasks for which we ordinarily use complex conceptual thought. In particular, it won't necessarily carve up the phenomena in a way that proves helpful for making moral judgments.

Think, for example, of the various ways in which we might divide people up into categories. One approach would be to develop concepts that did the best possible job of predicting and explaining behavior. (And here we might end up with concepts like *person with high serotonin levels*.) But the categories we construct using this approach may turn out to be not ideal when it comes time to make moral judgments. Indeed, it may turn out that the categories that prove most helpful in making moral judgments are completely orthogonal to the categories that prove most helpful in generating predictions and explanations.

Assuming that we do want to make moral judgments, it seems that we will need to develop additional, non-scientific concepts that help us to pick out the morally relevant categories. Ultimately, we will then be left with two different ways of carving up the same class of phenomena. We will have concepts that pick out the categories that prove most helpful in prediction and explanation (e.g., person with high serotonin levels) and also concepts that pick out the categories that prove most helpful in making moral judgments (e.g., morally good person). We will then need a complex system of rules that enables us to move from one set of concepts to the other.

For cognitively limited creatures like ourselves, this level of specialization might be a major problem. We would have to retain in our minds two distinct systems of concepts, two distinct kinds of psychological mechanisms, two distinct sets of propositional attitudes. Whenever we were engaged in tasks that involved both prediction and moral judgment, we would have to shift back and forth from one system of categories to the other. All this would impose a substantial demand on our cognitive resources.

In short, the sort of approach we now associate with science has both advantages and disadvantages. The chief advantage lies in its *predictive power*; the chief disadvantage lies in the resulting *conceptual complexity*.

5. There is, however, another possible approach. Instead of having one system of concepts for use in generating predictions and then a second, completely separate system of concepts for use in making moral judgments, we could have a single system of concepts that was used for both of these tasks. This single system of

concepts might not do a perfect job either at generating predictions or at making moral judgments, but it could do at least an adequate job of *both*. Hence, although this system of concepts might not afford us the greatest possible predictive power, it would do quite a bit to reduce the amount of cognitive complexity we needed to handle.

For an analogous case, consider the various ways we might come to conceptualize the weather. In thinking about the weather, there is a need to make *predictions* about what conditions will arise in the future, and there is also a need to make *evaluations* of whether these conditions are good or bad for certain purposes. What sorts of concepts would best enable us to achieve these goals? One approach would be to have a system of concepts that was specifically suited to the task of making predictions and then another, entirely separate system of concepts that was specifically suited to the task of evaluation. But such an approach might leave us with a large and unwieldy array of distinct ways of carving up the same class of phenomena. We might therefore be better served by a single system of concepts that wasn't ideally suited either for prediction or for evaluation but could serve us at least fairly well in both of these tasks.

It is certainly conceivable that folk psychology uses a system of concepts that works more or less along these lines. That is, it is conceivable that folk-psychological concepts are constructed in such a way that they do an adequate job at helping us both with prediction and with moral judgment, though perhaps without doing an absolutely ideal job in either of these two domains. What we want to know now is whether there are any general theoretical arguments against the view that folk psychology works in this way.

Thus far, we have been considering one possible argument. This argument relies on an analogy between folk psychology and systematic science. It points out that systematic scientific research programs typically *don't* try to develop a small set of concepts that enable us to do at least passably well at a wide variety of different tasks. Instead, they typically seek to develop concepts that enable us to do the best possible job at a specific range of tasks (prediction, explanation, etc.), even if they thereby end up coming up with concepts that aren't especially helpful in the task of making moral judgments. The argument then suggests that this fact about the concepts used in systematic science gives us reason to expect to find something similar in the concepts used in folk psychology.

At least for the sake of argument, we have been accepting all of the relevant claims about the nature of systematic science. The key question then becomes whether these claims can justify the relevant inferences about folk psychology.

6. But when the question is put in these terms, one notes immediately that folk theories are quite different from the sorts of theories one typically develops in systematic scientific research programs. Clearly, the two kinds of theories occupy two very different kinds of roles in our lives, and there is therefore little reason to expect that people look to them to fulfill the same needs. Most importantly for present purposes, it seems that people are far more reluctant to tolerate conceptual complexity in a folk theory than they are in the theories they employ in systematic research programs.

In systematic research programs, one can easily deal with the problem of conceptual complexity through a division of cognitive labor. No individual researcher needs to learn all of the scientific concepts; each only needs to know the concepts used in one particular domain of inquiry. Thus, science as a whole can acquire an extraordinary level of conceptual complexity even without any individual person grasping more than a tiny fraction of the total.

This solution is not available in the case of folk psychology. We cannot make do with a system in which one person only knows the emotion concepts, another only knows the trait concepts, and so on. We will only be able to do a tolerable job of getting around in the world if each person has some grasp of the whole of folk psychology. In fact, this seems to be one of the fundamental differences between folk theories and systematic research projects. We do not look to folk theories for a system that can serve, at least in principle, to generate perfectly accurate predictions. We look to them for tools that can help creatures like us – with all of our cognitive limitations – to accomplish certain practical goals.

Ultimately, then, it seems that we have good reason to expect that the concepts used in folk psychology will differ in certain respects from the concepts used in systematic research. In systematic research projects, one should expect to find an enormous array of different concepts, with each concept highly specialized for one particular use. But there is good reason to expect that folk theories will work somewhat differently. In a folk theory, one should expect to find concepts that are less highly specialized and can therefore be used in a wider variety of different tasks. Each concept might be specific to one particular domain of phenomena, but it will be constructed in such a way as to help us do almost anything we might want to do with the phenomena in that domain. Thus, instead of expecting to find a clear distinction between concepts used for prediction and concepts used for moral judgment, one should expect to find concepts that are not specialized for either of these two tasks but are constructed in such a way that they can do a decent job of both.

9.3. III

There seems to be a widely shared intuition that moral considerations just *couldn't* be playing any fundamental role in the basic concepts of folk psychology. Researchers who hold this intuition have not backed it up with systematic arguments. In fact, they have not even mentioned it explicitly. Yet the underlying intuition comes through quite clearly in the incredulous stares one receives whenever one suggests that some particular folk-psychological concept might be best understood as having moral features.

My concern here has been with the question as to whether there actually are any general theoretical arguments in favor of this intuitive view. I focused in particular on the argument that we have reason to expect that folk psychology will show certain fundamental similarities to scientific inquiry. This argument did not fare especially well on closer inspection. In fact, it seems that we actually have some reason to expect that folk psychology will differ from science in the relevant respects.

Of course, it is possible that there really are good arguments for the view that moral considerations can't play any fundamental role in folk-psychological concepts and that these arguments have simply eluded my grasp thus far. In that case, I would want to know exactly what the relevant arguments are. Clearly, we should not reject a hypothesis simply because it goes against our philosophical preconceptions. What we need now are definite theoretical proposals that generate testable predictions about the structure of people's folk-psychological concepts.

NOTES

[1] I am grateful to Daniel Hutto and Matthew Ratcliffe for comments on an earlier draft.

[2] For some contributions to this debate, see Adams and Steadman (2004a, b, forthcoming), Harman (2006), Malle (2006), McCann (2005), Meeks (2004), Mele (2003), Morton (2006), Nadelhoffer (2004, 2005), Nichols and Ulatowski (2006), Sverdlik (2004), Turner (2004), Yoo (2004), and Young et al. (2006).

[3] For the original experiment, see Knobe (2003a). For further replications and extensions, see Adams and Steadman (forthcoming), Knobe (2003b, 2004), Knobe and Mendlow (2004), Nadelhoffer (2005, 2006 forthcoming-b), Malle (2006), McCann (2005), and Nichols and Ulatowski (2005).

[4] See especially Bratman (2000), Copp (1995), Lewis (1989), Sayre-McCord and Smith (2005), and Smith (1994). As far as we know, the only previous empirical studies of people's use of the concept are the excellent experiments in Malle and Edmondson (2006).

[5] For evidence against particular alternative explanations, see Knobe (2004, forthcoming-a), Knobe and Mendlow (2004), and Young et al. (2006). Advocates of specific alternative explanations have also provided interesting and important evidence against competing alternative explanations (see especially Nadelhoffer forthcoming-b; Nichols and Ulatowski 2006). Note that all of this evidence is directed against explanations that were offered a number of years ago. More recent research has led to the construction of new alternative explanations that accommodate all of the existing data while still assigning no fundamental role to moral considerations in people's folk-psychological concepts (e.g., Malle 2006; Nichols and Ulatowski 2006). It will be interesting to see how these new explanations fare in accounting for the results of future experimental studies.

[6] The idea that folk psychology resembles a scientific theory was perhaps first developed by Churchland (1981) and then came to play an important role in 'theory of mind' research as a result of work by Gopnik and Wellman (1992), Gopnik and Meltzoff (1997), and others. Since that time, numerous researchers have argued that folk psychology does not use quite the same *methods* that we find in scientific inquiry, but almost all of these researchers have assumed that folk psychology should be understood in terms of the typical *goals* of science (prediction, explanation, etc.). In a number of recent works, however, this assumption has been called into question. See especially Andrews (2006), Hutto (2004), Morton (2003), and Wilkes (1981).

REFERENCES

Adams F, Steadman A (2004a) Intentional action in ordinary language: core concept or pragmatic understanding? Analysis 64:173–181.

Adams F, Steadman A (2004b) Intentional action and moral considerations: still pragmatic. Analysis 64:268–276.

Adams F, Steadman A (forthcoming) Folk concepts, surveys, and intentional action. Proceedings of the international conference "Intentionality, deliberation and autonomy – the action theoretic basis of practical philosophy," Sienna, Italy.

Andrews K (2006) The functions of folk psychology. Unpublished manuscript, York University.

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press, Cambridge, MA.
Bloom P (2004) Descartes' baby: how the science of child development explains what makes us human. Basic Books, New York.
Bratman M (2000) Valuing and the will. Philos Perspect: Action Freedom 14:249–265.
Churchland P (1981) Eliminative materialism and the propositional attitudes. J Philos 78:67–90.
Copp D (1995) Morality, normativity, and society. Oxford University Press, New York.
Faucher L, Mallon R, Nichols S, Nazer D, Ruby A, Stich S, Weinberg J (2002) The baby in the labcoat: why child development is an inadequate model for understanding the development of science. In: Carruthers P, Stich S, Siegal M (eds) The cognitive basis of science. Cambridge University Press, Cambridge, pp335–362.
Gopnik A, Meltzoff AN (1997) Words, thoughts, and theories. MIT Press, Cambridge, MA.
Gopnik A, Wellman H (1992) Why the child's theory of mind really *is* a theory. Mind Lang 7:145–171.
Gopnik A, Glymour C, Sobel D, Schulz L, Kushnir T, Danks D (2004) A theory of causal learning in children: causal maps and bayes-nets. Psychol Rev III(1); 1–31.
Harman G (2006) Intending, intention, intent, intentional action, and acting intentionally: comments on Knobe and Burra. J Cogn Culture 6: 269–276.
Hutto D (2004) The limits of spectatorial folk psychology. Mind Lang 19:548–573.
Keil F (1989) Concepts, kinds, and cognitive development. MIT Press, Cambridge, MA.
Knobe J (2003a) Intentional action and side effects in ordinary language. Analysis 63:190–193.
Knobe J (2003b) Intentional action in folk psychology: an experimental investigation. Philos Psychol 16:309–324.
Knobe J (2004) Intention, intentional action and moral considerations. Analysis 64:181–187.
Knobe J (2005a) Folk psychology and folk morality: response to critics. J Theor Philos Psychol 24:252–258.
Knobe J (2005b) Theory of mind and moral cognition: exploring the connections. Trends Cogn Sci 9:357–359.
Knobe J (2006) The concept of intentional action: a case study in the uses of folk psychology. Philos Stud 130: 203–231.
Knobe J, Burra A (2006) Intention and intentional action: a cross-cultural study. J Culture Cogn 6: 113–132.
Knobe J, Mendlow G (2004) The good, the bad and the blameworthy: understanding the role of evaluative reasoning in folk psychology. J Theor Philos Psychol 24:252–258.
Leslie A, Knobe J, Cohen A (2006) Acting intentionally and the side-effect effect: 'theory of mind' and moral judgment. Psychol Sci 17: 421–427.
Lewis DK (1989) Dispositional theories of value. Proc Aristotelian Soc Suppl 63:113–137.
Malle B (2006) Intentionality, morality, and their relationship in human judgement. J Cogn Culture 6: 87–113.
Malle BF, Edmondson E (2006) What are values? A folk-conceptual investigation. Unpublished manuscript, University of Oregon.
McCann H (2005) Intentional action and intending: recent empirical studies. Philos Psychol 18:737–748.
McCauley RN (2000) The naturalness of religion and the unnaturalness of science. In: Keil F, Wilson R (eds) Explanation and cognition. MIT Press, Cambridge, MA, pp61–85.
Meeks R (2004) Unintentionally biasing the data: reply to Knobe. J Theor Philos Psychol 24:220–223.
Mele A (2003) Intentional action: controversies, data, and core hypotheses. Philos Psychol 16:325–340.
Morton A (2003) The importance of being understood: folk psychology as ethics. Routledge, London.
Morton A (2006) But are they right? The prospects for empirical conceptology. J Cogn Culture 6: 159–163.
Nadelhoffer T (2004) Blame, badness, and intentional action: a reply to Knobe and Mendlow. J Theor Philos Psychol 24:259–269.
Nadelhoffer T (2005) Skill, luck, control, and intentional action. Philos Psychol 18:343–354.
Nadelhoffer T (2006) Bad acts, blameworthy agents, and intentional actions: some problems for jury impartiality. Philos Explorat 9:203–219.

Nadelhoffer T (forthcoming) On saving the simple view. Mind Lang.
Nichols S, Ulatowski J (2006) Intuitions and individual differences: the Knobe effect revisited. Unpublished manuscript, University of Arizona.
Pinker S (1997) How the mind works. Norton, New York.
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 4:515–526.
Sayre-McCord G, Smith M (2005) Desires…and beliefs…of one's own. Unpublished manuscript, University of North Carolina-Chapel Hill.
Sellars W (1956) Empiricism and the philosophy of mind. Minnesota Stud Philos Sci 1:253–329.
Smith M (1994) The moral problem. Blackwell, Oxford.
Stich S, Nichols S (1998) Theory theory to the max. Mind Lang 13:421–449.
Sverdlik S (2004) Intentionality and moral judgments in commonsense thought about action. J Theor Philos Psychol 24:224–236.
Turner J (2004) Folk intuitions, asymmetry, and intentional side effects. J Theor Philos Psychol 24:214–219.
Watson G (1975) Free agency. J Philos 72:205–220.
Watson G (1987) Free action and free will. Mind 96:145–172.
Wilkes K (1981) Functionalism, psychology, and the philosophy of mind. Philos Topics 12:1.
Yoo J (2004) Folk psychology and moral evaluation. J Theor Philos Psychol 24:237–251.
Young L, Cushman F, Adolphs R, Tranel D, Hauser M (2006) Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. J Cogn Culture 6: 291–304.

MARTIN KUSCH

10. FOLK PSYCHOLOGY AND FREEDOM OF THE WILL

10.1. INTRODUCTION

Philosophers discussing folk psychology tend to focus on some categories at the expense of others. Belief receives most attention; desire and action come in joint second place. The emotions feature only when questions of universality and cultural variation are addressed. Other mental states – e.g. intentions, volitions and moods – appear hardly at all. The one-sided diet has its drawbacks. One such drawback is a lack of engagement and exchange with other philosophical and scientific disciplines. Intentions, volitions and moods are objects of intense curiosity elsewhere, both inside and outside of philosophy: intentions are investigated in the philosophy of action and the law; moods (like *Angst*) have attracted the scrutiny of existentialists and psychiatrists; and volitions have been central to the millennia-old debate over freedom of the will. I believe that the debate over folk psychology would be richer and deeper if it made contact with the work in these areas. The same is true in the opposite direction: The philosophies of action, moods and the will might well profit from reflecting on the controversy over the status of folk psychology.

In this paper, I shall try to bring theorising about folk psychology into contact with philosophical and sociological work on freedom of the will. In order to allow for a reasonable degree of argumentative resolution, I shall concentrate on just two kindred proposals, one each from the two respective discussions. The free-will literature will be represented by a recent book, *Understanding Agency* (2000) by the social theorist Barry Barnes. I shall compare and contrast Barnes' important work on freedom of the will with my own so-called sociophilosophy of folk psychology (Kusch 1999). By "sociophilosophy" I mean a philosophy that takes its starting point from the results of, and a critical engagement with, the social sciences in general, and the sociology of knowledge in particular. Although Barnes does not use the term "sociophilosophy" himself, it will be obvious to anyone familiar with Barnes' wide-ranging oeuvre that he sympathises with the programme behind the term.

10.2. THE FOLK-PSYCHOLOGY DEBATE

The past 25 years have seen a vigorous debate – first and foremost in philosophy – over the nature, innateness, truth value, scientific promise and eliminability of folk psychology. In this debate, folk psychology has usually been understood as centred around the concepts of *belief*, *desire* and *action*. The controversy first developed around the idea that folk psychology is a folk *theory*. This view of folk psychology is usually called "the theory theory of folk psychology" (Morton 1970). Like any other theory, psychological folk theory allegedly is a network of laws and enables

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 175–188.

its users to explain and predict observable behaviours by postulating unobservable "theoretical entities". In the case of folk psychology, these laws are platitudes like "someone who is thirsty desires to drink", or "someone who believes herself successful is likely to feel pride in her achievement". The theoretical entities are, first and foremost, beliefs and desires.

As with any scientific theory, so also in the case of folk psychology one might wonder whether we should be realists or instrumentalists about the theoretical entities in question. And indeed, one strand of the debate has concerned precisely this point. More controversial has been the thought that folk psychology would – at least eventually – have to walk the plank with other folk theories: just like folk physics or folk biology have been replaced by scientific physics and scientific biology, so folk psychology would eventually be "eliminated" in favour of neuroscientific theories (Churchland 1989; Stich 1983). "Eliminativists" argue that we already now have good reason to believe our current folk psychology to be scientifically unsatisfactory. For instance, it provides no explanation for phenomena like sleep, motor control, or mental illness. In arguing for eliminability, the foes of folk psychology assume the "descriptive theory of reference". That is to say, they assume that the theoretical entities of a theory share its fate. In a slogan, if folk psychology turns out to be false, then beliefs, desires and actions do not exist.

Neither the assimilation of folk psychology to a scientific theory nor the arguments for elimination have met with universal approval – to put it mildly. The friends of folk psychology have argued that the platitudes of folk psychology are *norms* rather than *laws* of nature (Dennett 1993); that folk psychology does not primarily serve to explain and predict (Wilkes 1984); that *belief* and *desire* are observational rather than theoretical terms (Double 1985); that our ability to predict and understand others' behaviour is not grounded in our possession of a folk theory but based on a capacity to model and simulate what others think and feel (Goldman 1995; Gordon 1995; Harris 1991; Heal 1995); that psychological folk theory is innate, modular and not eradicable (Clark 1987; Fodor 1992; Segal 1996); and that it is the backbone of a highly successful scientific psychology (Horgan and Woodward 1991).

10.3. THE SOCIOPHILOSOPHY OF FOLK-PSYCHOLOGY

In my book *Psychological Knowledge: A Social History and Philosophy* (1999), I have tried to introduce a new perspective by developing a "sociophilosophy of folk psychology", that is, by working out the consequences for folk psychology of thinking of ourselves as "highly gregarious and deeply interdependent social beings" – to borrow a felicitous phrase from Barnes (2000, p. ix). This project has both a critical and a constructive side. The critical side consists in showing that all sides of the folk-psychology debate work with radically impoverished conceptions of our profound sociability. Most writers on folk psychology model the folk psychologist – whether she be an infant or an adult – on a socially isolated, individual natural scientist who tries to interpret some non-social phenomenon. Communicative

interaction between the one who interprets and the one who is being interpreted is neglected. And ignored is the fact that folk psychological interpretations are shared and discussed with others. There is something deeply individualistic also about the notion that we could decide to throw folk-psychological old-speak overboard. Clearly the abandonment of folk psychology would be a very complex social process, so much so that one might wonder whether it is at all *socially* possible.

The constructive side of my sociophilosophy consists of the proposal to think of folk psychology as our most fundamental social institution. By "social institution" I mean a self-referring and self-validating system of talk and behaviour (cf. Barnes 1983, 1988; Searle 1995). For instance, money is money because we talk of it as money and behave towards it as money. Money is what we collectively take to be money. Moreover, social institutions are ways of coordinating behaviour; this makes them collective goods. Social institutions produce their characteristic artefacts – coins, for example; and they are only rarely universal. Finally, statements about social institutions cut across the distinction between the normative and the descriptive. "Promises are kept" can be both a description of a successful social institution of promise-keeping, *and* a normative statement by means of which we remind one another to stick to this institution.

One of my central ways of arguing for my "institution theory of folk psychology" is modelled on Aristotle's method for demonstrating the superiority of his version of metaphysics over those of his predecessors: the method of "saving the phenomena". That is to say, I seek to show that *if* folk psychology were a social institution, then we would be able to explain why different authors have such differing intuitions about its nature. For instance, we could understand why some authors think that folk psychological platitudes are like descriptive laws of nature, whereas others are equally adamant that these platitudes are norms. As my example of the institution of promise keeping suggests, if folk psychology is a social institution, then the platitudes can be both. I also seek to show that folk psychological platitudes play a key role in keeping us predictable, comprehensible, accountable and susceptible to influence by others. So far from being a theory to be tested for accuracy and truth, folk psychology often plays the role of a "paradigmatic" theory in Kuhn's (1962) sense: It is us folk psychologists who have to prove our mettle by accounting for our actions in terms of folk psychological platitudes. It is us who are tested, not folk psychology as a theory. Finally, I marshal evidence for substantial cross-cultural variation regarding even central categories like *belief*, and I suggest that central aspects of our psychological phenomenology are artefacts of our folk psychology: our experience of our mental states follows the folk psychological platitudes, not *vice versa*.

10.4. FOLK PSYCHOLOGY AND FREEDOM OF THE WILL

The concepts of "the will" or "the willing subject" have played no great role in the folk psychology debate. I suppose the reason for this absence is straightforward: Analysts of folk psychology simply assume that any general discussion of folk

psychological categories covers the will as well and that the category of the will poses no special and distinct problem.

Even though the will has been largely absent from the folk psychology debate, it has, since antiquity, been vigorously discussed by metaphysicians and moral philosophers. I am thinking here of course of the controversy over the freedom of the will. There are noteworthy parallels and differences between this controversy and the folk psychology debate.

The first, and somewhat trivial, commonality is that in both cases we are dealing with questions regarding existence. The folk psychology debate concerns (at least in good measure) the question whether beliefs and desires exist; the controversy over the will concerns the question whether there is such a thing as a free will.

Moreover, in both disputes we face a conflict between our everyday discourse and scientific work. In our everyday discourse, we treat beliefs and desires as real entities, and we happily call at least some actions free or freely chosen. In the case of belief and desire, the science conflicting most strongly with common sense assumptions is neuroscience. Allegedly neuroscience does not need beliefs and desires to account for human behaviour; and beliefs and desires cannot be "smoothly reduced" to the theoretical entities of neuroscience. I have yet to see an argument in the free-will controversy that puts the case against free will in analogous terms, but at least some authors gesture vaguely towards the possibility that some results in brain physiology might enable one to put things in this way (e.g. Libet et al. 1999).

In the case of the controversy over freedom of the will, the relevant science has traditionally been physics. It has often been thought that the assumption of universal physical determinism causes problems for the possibility of free will. "Incompatibilists" maintain that such determinism is, well, *incompatible* with freedom of the will. "Pessimistic incompatibilists", or "hard determinists", conclude that freedom of the will is impossible and a mere illusion; "optimistic incompatibilists", or "libertarians", believe that all is well since the assumption of universal physical determinism is false. "Compatibilists", or "soft determinists" think that determinism and free will can happily coexist (Fig. 1).

The controversy over the freedom of the will is much older than the debate over beliefs. This is because the freedom of the will can be threatened – or be perceived to be threatened – not just by science, but also by theology or metaphysics.

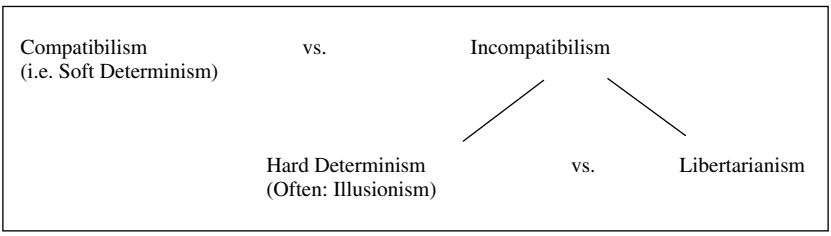

Figure 1. Positions on the freedom of the will

Universal determinism need not come in a physical garb, and divine foreknowledge or providence can seem to threaten free will no less than do physical principles.

The most important difference between the two disputes is the following. Whereas someone who denies the existence of beliefs rejects folk psychology as a whole, someone who opposes the idea that our wills are free does not thereby discard folk psychology rock, stock and barrel. This difference is already visible in the central question of each of the two controversies: One is: Do any of the categories of folk psychology refer to anything? And the other asks: Is the will free? The second question presupposes that the will exists. In other words, the free will debate takes much of our folk psychology for granted. Writers on the freedom of the will test our folk psychology only to a limited extent: They focus on those folk psychological platitudes that concern the will. And they do so by relating these platitudes to scientific results, to metaphysical assumptions, to our phenomenology or to general criteria of consistency.

10.5. BARNES' PROPOSAL

What would it mean to develop a sociophilosophy of free will? In broad outline, it would mean to propose a theory of the will that appropriately reckons with our profound sociability. This can be done in two ways. The *quick way* is to suggest that our discourse regarding the will is simply part and parcel of our folk psychology in general, and that whatever is true for the general case must also apply to the particular instance. Although this is true, it is not particularly helpful; it does not throw any new light on either the will or the folk psychology. And thus I prefer the *slow way* of bringing the will into the realm of sociophilosophy. The slow way too is premised on the idea that we are highly gregarious and deeply interdependent social beings. But the slow way seeks to connect this idea much more directly with our talk about freedom and the will. The slow way seeks to understand what role voluntaristic talk plays for social beings like us. It is here that Barnes' *Understanding Agency* has opened a new direction of research. In this section I shall reconstruct his somewhat complex argument in my own terms.

The starting point is observations concerning the distinction between actions that are done voluntarily and actions that are done involuntarily. Dictionary definitions suggest that we see a close connection between voluntary actions, freedom from external or internal coercion, rationality and responsibility. Voluntary actions are actions performed by rational agents who have made their own independent and free decisions. Sometimes we are willing to go so far as to say that voluntary agents are the uncaused causes of their actions. They are starting points of causal chains, but their decisions are not themselves caused. To explain a voluntary action is to appeal to reasons rather than causes. Involuntary actions are actions that the agent cannot help doing, actions that the agent was forced to do. Involuntary actions are caused and coerced actions. And to explain an involuntary action is to marshal causes rather than reasons.

Although we sometimes put great emphasis on the distinction between caused and uncaused actions, in other contexts we happily ignore it. For instance, we have no difficulties holding people responsible for actions that are not voluntary in the strong sense of being uncaused. We punish criminals even though we recognise that the deprivations of their childhood were partial causes of their decisions to commit crimes, and we continue to speak of voluntary actions despite the fact that we learn more and more about the myriad ways in which nature and nurture shape our thinking and acting.

There must then be two kinds of criteria at work in our thinking about voluntary actions. On the one hand, we have *strict criteria* for voluntary actions, criteria that demand that voluntary actions are the product of an uncaused causer, of a metaphysically free will. On the other hand, we also seem to admit more diffuse and *pragmatic criteria*, criteria that allow that voluntary actions and free decisions are the products of all sorts of causal intervention. This duality of criteria raises two questions: What, if anything, underlies the pragmatic set of criteria? And why do we have these two different ways of thinking about free actions?

To answer the first question (concerning the unity underlying the pragmatic criteria), it helps to consider the issue of what kinds of action and decision classifications must be important to highly gregarious and socially interdependent beings like us. A little reflection shows that such creatures must find it important to distinguish between two kinds of decisions by others: Decisions by others they can influence by means of verbal interventions; and decisions by others that they cannot so influence. I can bring about your decision to fetch me a glass of water by asking for it or by explaining how important the water is to me. "Fetching a glass of water" is a type of action that typically can be brought about by symbolic intervention. However, if you are pathologically obsessive about washing your hands, then typically I cannot stop you from doing so by asking you to stop. Your decision to wash your hand thus falls on the other side of the divide.

Barnes suggests that we see our distinctions between voluntary and involuntary actions, and between free and coerced decisions, in light of the distinction between decisions that can be influenced by communication, and decisions that cannot be so influenced. What unifies our set of pragmatic criteria for attributing free decisions and voluntary actions is the idea of being susceptible to verbal intervention. A careful look at the decisions that we end up classifying as either free or coerced suggests that susceptibility to change on the basis of symbolic intervention is central. Put differently, it seems that for us a decision is free if it "could have been otherwise if symbolic intervention had occurred" (Barnes 2000, p. 73). If this is correct, then as "folk metaphysicians" we often are compatibilists concerning causality and freedom: What makes a decision free is not that it is uncaused but that it is not causally insulated from one specific kind of cause: verbal intervention.

If all this is true, why is it that we do not spontaneously produce "susceptibility to change through communication" as the decisive criterion for free actions? And why do we have such strong intuitions concerning the importance, for voluntary decisions, of rationality, responsibility and uncaused causes?

To find the answers to these questions, we have to remain firmly focused on our profound sociability – the fact that we constantly seek to influence each other through interaction in general and communication in particular. Let us call "causal discourse" the modes of speech in which we centrally rely on the language of "causes" and "coercion"; and let us reserve "voluntaristic discourse" for modes of speech in which we employ the categories of "responsibility", "rationality", "free action", or "free decision". Now, which of these discourses is central when we seek to influence others to co-operate with us and to act in ways we would like them to act? Obviously, the answer must be "voluntaristic discourse". The best way for you to convince me to help you make dinner, is to appeal to my sense of honour and freedom. You might be able to coerce me – by literally twisting my arm – but coercion is a precarious tool that easily destabilises and endangers social relationships. Almost always we therefore prefer to act in more subtle ways. We constantly hint and imply that we take each other to be rational and responsible uncaused causes of our actions, and we do so while suggesting that the best way to remain with this status is to yield to our demands. You get me to help you with dinner by saying and implying that helping you is precisely the sort of action that a radically free and uncoerced rational agent would engage in. Of course, in yielding to your demand I am acting freely by our pragmatic – rather than by our strict – criteria.

Here then lies the key to understanding why we do not openly abide by our pragmatic criteria: The strict (libertarian) criteria are a necessary illusion – an illusion that is absolutely central to the ways in which we highly gregarious and deeply interdependent social beings go about influencing one another. Operating the strict criteria for freedom is central to the discourse in which we seek to influence others. In doing so we try to increase the domain of decisions that are free in the pragmatic sense.

10.6. BARNES' PROPOSAL AND THE SOCIOPHILOSOPHY OF FOLK PSYCHOLOGY

Barnes' sociological account of freedom of the will has obvious parallels with my sociophilosophy of folk psychology. We both offer social accounts of phenomena that (much of) the philosophical tradition has conceived of primarily in individualistic terms: folk psychology has been treated as a theory in and of the individual mind; free will has been looked upon as the possession of the individual being. Barnes' theory also makes the move I call "Aristotelian": Just as the sociophilosophy of folk psychology tries to explain how it comes about that philosophers construe folk psychology in individualistic terms, so Barnes' theory shows why we end up thinking about voluntary action and free will according to strict criteria. One might say that the reflections about freedom of the will of at least some philosophers – libertarians – have been led astray by our everyday talk about freedom of the will. Starting from our spontaneously produced intuitions about free will, libertarians have tried to build theories that do justice to these intuitions. But they

have failed to see that these intuitions are an illusion produced by our central social institution for pressing others into our service.

Barnes' and my proposals are also related, of course, in that we both make the concept of social institution central to our respective theories. I speak of folk psychology as our most fundamental social institutions; Barnes treats the responsible agent and voluntaristic discourse in the same way. Folk psychology and voluntaristic discourse are not of course independent social institutions. Instead, it seems most natural to treat voluntaristic discourse as a central part of our folk psychology in general. As such a central part, talk of willing and choosing is inseparable from other core concepts of our folk psychology, such as belief or desire. After all, we assume that rational decisions should be based upon true beliefs and reasonable desires.

Finally, seeing folk psychology, and voluntaristic discourse within it, as our most fundamental institution helps to understand why neither is likely to disappear any time soon. Social life is built around these institutions, and thus they are presupposed in almost everything we do. Any attempt to argue about the nature of folk psychology must presuppose its terms; and any attempt to persuade others to give up the notion of free will inadvertently appeals to their self-understanding as free and rational agent.

10.7. BARNES AND THE CONFLICT OF INTUITIONS

Perhaps the most intriguing aspect of Barnes' book is that it provides a sociological explanation for why the dispute over the freedom of the will has proven so difficult – nay, so impossible! – to resolve. In order to highlight this point, it might help to consider Stephen Schiffer's pessimism regarding all solutions to what he calls the "Paradox of Free Will" (Schiffer 2002). The paradox is generated by accepting all three of the following:

[1] We have free will; at least some of the things we do we do freely, of our own free will.

[2] Everything we do is such that we were caused to do it by factors over which we had no control, perhaps factors that obtained even before we were born.

[3] If [1] is false, [2] is true.

Different philosophical positions regarding the freedom of the will resolve the paradox by denying different premises: compatibilism rejects [3], libertarianism [2] and hard determinism [1]. Let us say that a paradox has a "happy-face solution" if and only if we can escape it by rejecting one of its premises (Schiffer 2002). Compatibilism, libertarianism and hard determinism all believe that the Paradox of Free Will has a happy-face solution. Schiffer disagrees. As he sees it, none of the three options has a satisfactory account of the initial plausibility of its chosen "odd-guy-out". On Schiffer's analysis, there is a fundamental tension in our concept of free will, a tension that expresses itself in our acceptance of the incompatible propositions [1]–[3]. Needless to say, not many philosophers writing on freedom of the will are likely to accept Schiffer's claim – to properly convince them one

would have to enter the debate between the three camps and show that all three lack a satisfactory account. This is no small task, and here is not the place to take it on. Fortunately, for present purposes I do not need to show that Schiffer is right; all three sides in the debate over the freedom of the will can agree that we have these (three) fundamentally conflicting intuitions regarding free will and determinism, and that we can bring these intuitions into harmony only by finding a way to dispose of one of them.

Putting things in these terms should immediately stimulate curiosity concerning the question why we have such conflicting intuitions in the first place. And it is just at this point that Barnes' speculative sociological proposal is an intriguing and provocative source of insight. Given only some limited assumptions about what it is to be a communicating and interacting human being, Barnes provides a sketch of why such beings will end up with something of a bifurcated conception of free will and responsibility.

10.8. BARNES' COMPATIBILISM

Does Barnes' *sociological* analysis provide support for a particular *philosophical* view of freedom of the will? Barnes thinks it does: He maintains that his analysis supports compatibilism. While there is a certain sense in which I agree with this claim, I disagree with Barnes' reasoning in its support.

First of all, it must be noted that *Understanding Agency* uses the term "compatibilism" for phenomena of different levels and kinds. In the interest of clarity, Barnes' different uses of "compatibilism" can be distinguished as follows. Compatibilism$_1$ is a descriptive claim about our everyday intuitions and ways of talking; according to compatibilism$_1$ many of our everyday intuitions are compatibilist (in the sense of allowing that at least some types of actions can be both free and causally determined). Compatibilism$_2$ is a descriptive claim about the relations between causal and voluntaristic, libertarian and compatibilist discourses in our social world; compatibilism$_2$ tells us that causal and voluntaristic discourses can peacefully coexist but that libertarian discourse is due to an illusion. Finally, compatibilism$_3$ is the normative-prescriptive view that we should be compatibilists in the everyday sense (we should allow that at least some actions can be both free and causally determined).

What support does Barnes offer for compatibilism$_3$? One line of argument starts from the observation that in everyday life we sometimes talk and judge like compatibilists, and that sometimes we speak and think like libertarians. But although both positions appear in everyday life, this does not make them equal. Barnes calls our everyday compatibilism "robust", "easy-going" and "frequent", and he speaks of our occasional libertarianism as "unsatisfactory" (Barnes 2000, p, 4, 111). Barnes feels that this unsatisfactory character of our libertarianism is the most apparent in legal contexts: Whatever libertarianism "is to be found in our practical life is typically sustained in a formally unsatisfactory form, as is the case, for example, in many legal contexts" (Barnes 2000, p. 114). Barnes gives one example. In many

legal settings it is assumed that choices are in principle uncaused – that they are free in a libertarian sense. At the same time, lawyers and judges have to reckon with the possibility that various causes – from upbringing to drugs – might impinge upon the choice. The standard 'solution' is to suggest that causal circumstances can, to an extent, constrain or restrict the free will. Moreover, in any given case, the degree of deserved blame depends on the extent to which the free will was constrained. What makes this legal practice "unsatisfactory" is that "sadly …measurement of the relevant 'constraints' is still a little difficult" (Barnes 2000). In other words, it is close to impossible to make sense of the required "degrees of freedom".

These passages give the impression that Barnes wishes to defend compatibilism$_3$ with reference to compatibilism$_1$. Unfortunately, Barnes' critique of everyday libertarianism is not decisive. The problem of defining degrees of freedom is real enough, but it is not peculiar to libertarianism. All voluntaristic discourses face the problem of having to declare agents not just responsible or lacking in responsibility *simpliciter*, but of having to declare agents more or less responsible. And since the concepts of responsibility and freedom are inseparable, degrees of responsibility call for degrees of freedom. If Barnes wants to tip the balance in the compatibilist's favour, then he must show us a way in which the compatibilist can avoid these "formal" problems. And this he fails to do.

In other places Barnes seems to back compatibilism$_3$ by invoking compatibilism$_2$. Allegedly, sociological investigation shows that our libertarian intuitions are illusions created by our participation in certain forms of discourse. And this indirectly strengthens the case for compatibilism$_3$. But here too we must tread carefully. In particular, we must avoid committing the genetic fallacy. It is a mistake to think that certain ideas are false just because they frequently are triggered either by our manipulations of others' self-image or by our inability to think sociologically. It may well be true that libertarian intuitions often arise in suspect circumstances. However, pointing this out leaves open the possibility that we might yet find convincing justifications for them – despite their doubtful origins.

10.9. BARNES' ILLUSIONISM

It seems to me that Barnes' liberal use of the category of illusion quite generally undermines his case for compatibilism$_3$. Barnes' repeated claims that we do not understand our own voluntaristic discourse and its functions makes it difficult to distinguish his position from hard determinism – after all, illusionism and hard determinism tend to go together. Barnes holds that our actions are always caused and always caused deterministically. He also insists that central elements of our voluntaristic discourse are based on illusions: We social agents are blind to the ways in which we are susceptible to the influences of others; we fail to see what the formula "could have done otherwise" really amounts to (it is tantamount to "could have been influenced by communication"); we are wrong (at least sometimes) in thinking that chosen actions cannot have causes; and we are confused in assuming that actions can be explained in terms that are central to our voluntaristic discourse.

Barnes commits himself to the last-mentioned view in speaking about the attempts of geneticists and biotechnologists to bring human behaviour into the scope of their theories: "…it can be perfectly natural and forgivable for them to consider that their accounts face competition from 'voluntaristic explanations'" (Barnes 2000, p. 118). The implication is that geneticists and biotechnologists should learn to know better. Finally, and especially tellingly, Barnes claims that we are all blind to the "projectivist" – rather than a "realist" – metaphysics of our voluntaristic discourse. Again it is legal contexts that are marshalled as evidence. Legal contexts are said to show that the interest in assigning a specific status – say 'guilty' – drives the attribution of mental states (like free or not free), and not *vice versa*:

> The move that is supposed to be made is from state to status, but it is clear that strong back-pressure, to say the least, exists from status to state. Prior concerns with status can and at times do shape and structure attributions of state …The most immediately obvious features of courtroom decisions actually raise the question of whether concepts like 'responsibility', 'choice', 'free will', 'agency' and so forth might not be secondary features of the institution of responsible action, mere rationalising accompaniments of procedures moved by pragmatic expediency (Barnes 2000, p. 14).

The illusionism in all this is obvious. And it is hard to see what remains of freedom of the will – even a compatibilist version – once we adopt this illusionism. What remains is that in our voluntaristic discourse we are often willing to talk and think along compatibilist lines. But this does not give us a prima facie case for compatibilism. Barnes' general illusionism undermines all respect for the actors' categories and untutored philosophical preferences. One cannot dismiss so much of the actors' perspective as illusion and yet see it as giving us a prima facie case for compatibilism. If our everyday thinking about free will and determinism is beset by so many different forms of misunderstanding and confusion, why then should we philosophical and sociological analysts take our everyday compatibilism seriously? If ordinary folks are wrong about nearly all other aspects of freedom and determination, why should we be right when talking and thinking in compatibilist ways?

10.10. ELIMINATIVISM AND CONSTRUCTIVISM IN THE DEBATES OVER FOLK PSYCHOLOGY AND THE FREEDOM OF THE WILL

Contrary to Barnes' *official* view – that his study supports compatibilism – it seems that his *real* message is hard determinism, or eliminativism, regarding the freedom of the will. On this latter reading Barnes' argument comes down to the following thought: Since we have conflicting intuitions and idioms for talking about free will, and since the status of the free and responsible agent is frequently used as a tool in the social manipulation of others, the expression "free will" does not refer to anything. Let me call this position "Barnes' unwanted eliminativist thought".

Barnes' "unwanted eliminativist thought" concerning freedom of the will is reminiscent of a well-known eliminativist argument regarding the folk-psychological category of belief. In his "empirical-philosophical" book *Belief, Language and Experience* (1972), the anthropologist Rodney Needham demonstrates meticulously that "belief" is neither a universal nor (for us) a univocal category. First, Needham

shows that a "bewildering variety of senses" attach to words in foreign languages that have been used as translations for the English verb "to believe". Second, Needham seeks to substantiate Evans-Pritchard's notorious view according to which the Nuer do not have "our" concept of belief. Third, Needham documents the etymology of the English word "belief" and the changes in the Western concept of belief. And fourth, he shows that the word "belief" has a plethora of different uses in modern English, uses that cannot be reduced to one specific core. Needham concludes that beliefs do not exist as "natural resemblances" amongst humans, and that "when people are said, without qualification, to 'believe' anything, it must be entirely unclear what kind of idea or state of mind is being ascribed to them" (Needham 1972, p. 188). Like Barnes for the concept of "the free will", so also Needham for the concept of "belief" emphasises the importance of institutions and norms: "Belief-propositions implicate, in part at least, jural features of the social world – statuses, moral obligations, norms of co-operation, etc. – and this means that sociological analysis is required if the propositions are to be correctly construed." (Needham 1972, p. 168.) Stephen Stich warmly applauds Needham's eliminativist conclusion: "… Needham's painstaking Wittgensteinian survey of belief in various cultures came to just the right conclusion: There is no such thing as believing that p" (Needham 1983, p. 226).

I am not pointing out the parallel between Barnes and Needham in order to lend support to either view. Far from it. I want to suggest that both the criticism that has been directed at Needham, and the alternative gloss that has been put on his empirical work, can be applied, *mutatis mutandis*, also to Barnes' eliminativism concerning the free will. Needham – and Stich – thinks that cultural variation in the category of belief shows that no one has beliefs. However, to reason in this way is to overlook a third alternative *between* universalist realism and eliminativism: This third alternative is constructivism. According to constructivism, (different types of) beliefs (and related mental states) are cultural artefacts. Rom Harré puts the point in this way:

> Needham's [conclusion] is still universal. No one has beliefs. If we consider Needham's reasons for drawing his conclusion, they are all reasons for taking belief to be a non-universal psychological property, a much more interesting hypothesis.… It seems to me that the correct conclusion to draw… is that belief is a mental state, a grounded disposition, but is confined to people who have certain social institutions and practices (Harré 1981, p. 82).

My sociophilosophy of folk psychology is an attempt to develop and defend this form of constructivism in a systematic fashion.

What would it mean to apply this form of constructivism to the freedom of the will? Here I can offer no more than a sketch of an answer. First, constructivism would amount to the claim that in different contexts we have – and other cultures might well have – different concepts of freedom of the will. Second, these concepts, these ways of conceptualising choices and deliberations, are partly constitutive of what they are about. Acts of choosing freely are *real* artefacts – not just *mere* artefacts, not mere illusions – of our culture and our institutional context. And third, there is no viewpoint from which one conceptualisation could be judged superior

with respect to others. Such judgements or viewpoints would make sense only if we could separate the conceptualisation from the phenomenon conceptualised.

Is this a plausible position? And how does it relate to the traditional alternatives in the free-will debate? It seems that the constructivist position regarding free will is plausible only on the assumption of some form of compatibilism. The idea must be that different forms or versions of the freedom of the will differ with respect to the question which types of causal determination imply coercion. For instance, a choice influenced by strong social pressure might be regarded as coerced in one context or culture, while the same kind of choice might be looked upon as free in a different setting. And the two different settings do not only shape how the choice is conceptualised and evaluated, they also influence how members of the respective cultures experience their situation. In other words, the phenomenology of freedom follows the culture-specific classification of actions.

Free-will constructivism does not cohere well with either libertarianism or hard determinism: The fit with libertarianism is bad, since a culture-specific classification cannot make it so that there are gaps in the causal nexus. And hard determinism turns all talk of freedom – constructivist or not – into an illusion. Hard determinism is forced upon us only on the assumption that freedom and causal determination can never go together. If Barnes' observations are correct, then this assumption is very foreign to our everyday talk and intuitions.

10.11. CONCLUSION

If the – admittedly sketchy – argument of Section 10.10 is at least roughly on target, then there is, after all, a route from Barnes' sociological study of our causal and voluntaristic discourses to a defence of some kind of compatibilism. But this defence now proceeds in a way not anticipated in *Understanding Agency*. The key to finding this new way lay in the debate over folk psychology, more precisely in constructivist reinterpretations of Needham's data concerning the variety and history of the concept of belief and similar concepts (in other cultures). Contextual variation concerning the concept of free will need not be taken as paving the way towards hard determinism (or eliminativism); it can also be read as supporting the notion that there are different *concepts* of freedom of the will, and hence different *freedoms* of the will.

I thus hope to have shown – at least relative to a small sample of positions – that a discussion across the divide between philosophy of folk psychology and philosophy of free will can be fruitful and illuminating for both sides. Philosophers of folk psychology can learn much from Barnes' sophisticated analysis of the role of certain folk-psychological concepts in our (strategic) interactions. Even more fundamentally, Barnes' relentless focus on our profound sociability provides a welcome corrective to the still dominant individualistic perspective of much philosophical psychology. But the enlightenment is not all one-sided: The folk-psychology debate too has produced suggestions and results that at least Barnes' theory of free will had better incorporate.

REFERENCES

Barnes B (1983) Social life as bootstrapped induction. Sociology 17:524–545.
Barnes B (1988) The nature of power. Polity Press, Cambridge.
Barnes B (2000) Understanding agency: social theory and responsible action. Sage, London.
Churchland PM (1989) Folk psychology and the explanation of human behaviour. In: Churchland PM(ed) A neurocomputational perspective. MIT Press, Cambridge, MA, pp. 111–127.
Clark A (1987) From folk psychology to naive psychology. Cogn Sci 11:139–154.
Dennett DC (1993) Three kinds of intentional psychology. In: Christensen SM, Turner DR (eds) Folk psychology and the philosophy of mind. Erlbaum, Hillsdale, NJ, pp. 121–143.
Double R (1985) The case against the case against belief. Mind 94:420–430.
Fodor J (1992) A theory of the child's theory of mind. Cognition 44:283–296.
Goldman A (1995) Interpretation psychologised. In: Davies M, Stone T (eds) (1995), pp. 74–99.
Gordon RM (1995) Folk psychology as simulation. In: Davies M, Stone T (eds) (1995), pp. 60–73.
Harré R (1981) Psychological variety. In: Heelas P, Lock A (eds) Indigenous psychologies: the anthropology of the self. Academic Press, London, pp. 79–103.
Harris PL (1991) The work of imagination. In Whiten A (ed) Natural theories of mind. Blackwell, Oxford, pp. 283–304.
Heal J (1995) Replication and functionalism. In: Davies M, Stone T (eds) (1995), pp. 45–59.
Horgan T, Woodward J (1991) Folk psychology is here to stay. In: Greenwood JD (ed) The future of folk psychology: intentionality and cognitive science. Cambridge University Press, Cambridge, pp. 149–175.
Kuhn T (1962) The structure of scientific revolutions. University of Chicago Press, Chicago.
Kusch M (1999) Psychological knowledge: a social history and philosophy. Routledge, London.
Libet B et al. (1999) The volitional brain: towards a neuroscience of free will. Imprint Academic, Thorverton.
Morton A (1970) Frames of mind. Oxford University Press, Oxford.
Needham R (1972) Belief, language, and experience. Blackwell, Oxford.
Schiffer S (2002) Skepticism and the vagaries of justified belief.http://www.umass.edu/philosophy/events/contextualism_conference_files /SchifferPaper.pdf.
Searle JR (1995) The construction of social reality. Allen Lane, London.
Segal G (1996) The modularity of theory of mind. In: Carruthers P, Smith PK (eds), Theories of theories of mind. Cambridge University Press, Cambridge, pp. 141–157.
Stich S (1983) From folk psychology to cognitive science. MIT Press, Cambridge, MA.
Wilkes K (1984) Pragmatics in science and theory in common sense. Inquiry 27: 339–361.

PART III

THE FRAGMENTATION OF FOLK PSYCHOLOGY

KRISTIN ANDREWS

11. CRITTER PSYCHOLOGY: ON THE POSSIBILITY OF NONHUMAN ANIMAL FOLK PSYCHOLOGY

11.1. INTRODUCTION

Humans have a folk psychology, without question. Paul Churchland used the term to describe "our commonsense conception of psychological phenomena" (Churchland 1981, p. 67), whatever that may be. When we ask the question whether animals have their own folk psychology, we're asking whether any other species has a commonsense conception of psychological phenomenon as well. Different versions of this question have been discussed over the past 25 years, but no clear answer has emerged. Perhaps one reason for this lack of progress is that we don't clearly understand the question. In asking whether animals have folk psychology, I hope to help clarify the concept of folk psychology itself, and in the process, to gain a greater understanding of the role of belief and desire attribution in human social interaction.

To start, we can construct a simple argument in favor of animal folk psychology, based on a standard definition of the term. According to what I am calling the standard view, humans attribute specific mental states to a target, using a folk psychological theory, a mental simulation, or some combination of the two in order to generate predictions of intentional behavior. To count as a proper target for folk psychological analysis, the agent must engage in behavior that is predictable through the attribution of beliefs and desires. We see this commitment in Daniel Dennett's intentional stance, for example (Dennett 1987, 1991).[1] That folk psychology centrally involves the attribution of beliefs and desires also seems to be endorsed by Alvin Goldman, who takes attribution of mental states via simulation to be the primary means for predicting behavior.[2] These views reflect the standard notion of folk psychology, according to which intentional agents predict others' intentional behaviors via the attribution of mental states. Though one doesn't have to use belief/desire attribution in every instance of predicting behavior, to have a folk psychology one must be able to conceive of others as the sorts of things that have beliefs and desires, and be able to use specific mental state attributions to predict behavior; thus you have to conceive of others as intentional agents.

On these views, folk psychology is not seen merely as a useful heuristic for making predictions, but rather it is thought that the *primary* function of folk psychology is the prediction of behavior.[3] Any other role for folk psychology, such as the explanation of intentional behavior, is derivative of prediction. This assumption is clearly made in discussions of animal and child theory of mind, and in the debates between simulation and theoretical accounts of the subpersonal

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 191–209.

mechanisms driving our folk psychological behaviors (Andrews 2003). Dennett accepts this position when he writes, "…our power to *interpret* the actions of others depends on our power…to predict them" (Dennett 1991, 29). Folk psychology as the attribution of beliefs and desires is presented as the simplest predictive heuristic available for making accurate enough predictions across different domains. For this reason, advocates of the predictive power of folk psychology suggest that we use the attribution of beliefs and desires when we make all sorts of predictions, from the prediction that someone will duck if you throw a brick at him (Dennett 1991) to the prediction that you will arrive on the 3 p.m. flight if you say you'll arrive on the 3 p.m. flight (Fodor 1989). If there were quicker or easier heuristics we could use to make accurate predictions of behavior across domains, then there would be no reason to think that these examples are in fact examples of folk psychological prediction (rather than the result of using some other heuristic device). Since the standard view of folk psychology promised to help us understand the nature of beliefs and desires by presenting them as things that are used to make fast, easy, and accurate enough predictions of behavior, we can define folk psychology as how (non-Laplacean) intentional agents routinely predict the behavior of other intentional agents, namely through the attribution of beliefs and desires. The claim here is that without a folk psychology, predictions of behavior would not get made routinely, easily, and accurately (enough) across domains. In a world without folk psychology, we would not be able to predict that someone would duck if a brick were thrown at him, nor would we be able to predict that you will arrive on the 3 p.m. flight given that you said you will arrive on the 3 p.m. flight. Instead, everyone would be 'baffling ciphers.' Given such views, we may be tempted to think that though there may be more complex ways of predicting intentional behavior, such as the Laplacean super-physicist's method of following a deterministic causal chain, there are no more efficient and accessible heuristics available to us than the attribution of beliefs and desires. If so, then any non-Laplacean who routinely, easily, and accurately predicts behaviors across different domains has a folk psychology.

This understanding of folk psychology allows us to construct a cheap argument for critter psychology:

Argument C

1. Any (non-Laplacean) intentional agent who routinely, easily, and accurately predicts the behavior of other intentional agents has the ability to attribute beliefs and desires, and thus has a folk psychology.
2. Animals are (non-Laplacean) intentional agents who routinely, easily, and accurately predict the behavior of their conspecifics, competitors, predators, and prey.
3. From (1) and (2) it follows that animals attribute beliefs and desires, and thus have a folk psychology.

Because the conclusion follows from the premises, and we can assume the truth of (2), any problem with this argument must rest with (1). The main problem with premise (1) is that we have good reason to think that there are methods other than the attribution of propositional attitudes for predicting behavior. Despite what others

may suggest, even without appeal to people's beliefs or desires we can predict that someone will duck when a brick is thrown at him, because that's just one thing people generally do; they move to avoid large flying objects. And we can predict that you will arrive at the airport at 3 p.m., because you said you would and people generally do what they say they'll do. Without the ability to attribute propositional attitudes we probably wouldn't understand *why* people do what they say they'll do, but we could still make predictions by generalizing over past behavior. Since we can make predictions of intentional agents using such a method, perhaps the animals can as well. This, I think, is enough to establish that the attribution of beliefs and desires should not be described as *the* method we use to predict behavior. That is, though we may sometimes appeal to beliefs and desires when predicting behavior, predicting doesn't begin and end with the attribution of mental states. There may be other mechanisms that undergird our ability to anticipate behavior.

This sort of view seems to be what Barbara Von Eckardt endorses, given her definition of minimal folk psychology as consisting of "(a) a set of attributive, explanatory, and predictive practices, and (b) a set of notions or concepts used in these practices" (Von Eckardt 1994, p. 300). Von Eckardt wants us to accept a wider conception of folk psychology that includes "any concept of generalization ordinary people use in their FP practices" (Von Eckardt 1994, p. 305), and her account leaves open the possibility that humans don't need to attribute mental states to make predictions of intentional behavior. I am sympathetic to Von Eckardt's account of folk psychology, and I think there is good reason to accept this critique of (1), for reasons I will present in Section 11.2. After we have in place a wider and, I believe, more satisfactory account of folk psychology, we can create a reconceptualized Argument C and examine what it might tell us about critter psychology.

11.2. HUMAN FOLK PSYCHOLOGY

The problems with the standard view seem to stem from two assumptions found in the various accounts of folk psychology. These assumptions are:

Homogeneity – All folk psychological practices are subsumed under one subpersonal mechanism, and that mechanism involves the attribution of beliefs and desires. For example, since prediction and explanation are both folk psychological practices, and they both involve the attribution of beliefs and desires, it is widely assumed that explanation and prediction are symmetrical; whenever we have generated a prediction of behavior, we also have an explanation of that behavior (Andrews 2003).

Narrowness of function – Prediction and explanation are the paradigmatic folk psychological practices. Other functions of folk psychology, such as coordinating, cooperating, justifying, bonding, evaluating, etc. are largely ignored or seen as variations on the prediction and explanation functions.

If we take as our starting point the idea that folk psychology is "our commonsense conception of psychological phenomena," then there is clear empirical evidence that our folk psychological practices are not homogeneous. Even if we limit ourselves to prediction, we find that not all predictions of behavior are generated by the

same mechanisms, as was hinted at Section 11.1. Instead, what has been found is that predictions are extremely sensitive to the context of the prediction, the kinds of knowledge available to the predictor, and the predictor's own biases and past experiences. For example, research in social psychology has found that if we have no information about people, we tend to assume they will act like we do (Kreuger 1998; Marks and Miller 1987; Mullen et al. 1985) and when we have just a little information, we will use stereotypes or base rates about how people in that kind of situation behave (e.g., Locksley et al. 1980, 1982). However, the better we know a person, the less we utilize these techniques. When we know someone fairly well, we can use inductive generalization over past behavior to predict that a person will continue to do in the future the same sorts of things we know she did in the past. And after observing someone's behavior for a time, we can generate personality traits that we attribute to that person, and form our expectations about that person's future behavior based on those traits (Ross and Nisbett 1991). For example, if we have decided that a person is generous, we will expect her to leave a good tip and to help out with extra departmental chores. While these techniques allow us to make good-enough predictions, it is widely thought that some of these techniques, especially trait attribution, have some limitations.[4]

This isn't to say that we never use belief/desire attribution in order to make predictions; at the personal level, at least, we do utilize this method as well. However, a concern arises about the degree of accuracy associated with this method of prediction. Humans are thought to be fairly accurate in their ability to predict behavior, yet there is reason to think that predictions using belief/desire attribution are not terribly accurate. If the belief/desire method of predicting behavior is inaccurate, and most of our predictions are accurate, there is good reason to think that we do not use belief/desire attribution for the majority of our predictions.

Research in social psychology suggests that by considering a person's beliefs and desires we might actually decrease the accuracy of our predictions, because considering someone's reasons for making a particular prediction causes us to look for confirming facts, not for defeaters (Wilson and LaFleur 1995). Research also indicates that people will judge a piece of predicted behavior to be more likely that they predict to be more likely if asked to provide reasons for their prediction. For example, a subject might think that an event has a 70% chance of happening, but after being asked to provide reasons for thinking the event will occur, she will come to think that its chances are greater than 70% (see Kunda 2002 for a discussion). When we predict what someone is going to do by attributing propositional attitudes, we are also coming up with reason explanations for someone to engage in that behavior. Thus, by virtue of considering a possible belief/desire set we automatically promote the behavior as more likely to occur. For example, suppose a student is trying to predict whether her professor will raise her mark if she comes to complain during office hours. The student might wonder about her professor's beliefs and desires, and decide that the professor has a dislike of conflict, and so will give in to her request very easily. In considering whether the professor really has this attitude, the student uses only a positive test and searches for evidence that the hypothesis

is correct, and comes up with a number of anecdotes suggesting this is so (the professor has avoided conflict in another situation, the professor said he is a pacifist, etc.). The student is overconfident about the likelihood of her prediction. The act of trying to determine a person's mental state leads the student to talk herself into the theory she comes up with. This is a biased method of seeking evidence and does not result in accurate predictions.

The attribution of propositional attitudes doesn't exhaust the adult human's methods of predicting behavior. But further, given that our predictions are generally accurate and that research suggests our predictions based on belief/desire attribution may often be inaccurate, there is reason to suspect that it may not be nearly as common a method of predicting behavior as is suggested by the standard view. Humans are clearly pluralistic in the techniques underlying their predictive behaviors, and thus the homogeneity assumption is false with regard to prediction. There is also evidence that we use different techniques to explain different kinds of behavior and different people's behavior (Malle 2004). As social psychologists learn more about the mechanisms humans use while practicing folk psychology, the evidence against the homogeneity assumption increases.

Turning to the second assumption, the narrowness of the function of folk psychology, I want to suggest that folk psychology should be seen as more than predicting, and derivatively, explaining behavior. Narrowness is a problem because, coupled with the homogeneity assumption, it results in hasty generalizations about the mechanisms underlying other folk psychological practices. If we begin with an understanding of folk psychology as how the folk understand psychological phenomena, it seems apparent that such an understanding can take many forms. I can understand in the *Verstehen* sense, insofar as your behavior makes sense to me and I can relate to it, even if I don't have a covering law that describes the behavior. I can understand by feeling empathy for you. Or I can understand you by coordinating my behavior with yours. I can understand you insofar as your behaviors don't surprise me. I can understand you by knowing what you mean. I can understand you when I am able to explain your actions to someone else. I can understand you by justifying your behavior. In short, our folk psychological understanding involves the class of social behaviors that make up our social environment. By focusing on only one of those behaviors, we ignore the richness and the complexity of human social cognition.

What are we doing when we exercise our folk psychological skills? Fundamentally, we are interacting with others as intentional agents, even if that interaction takes the form of *analysis*. For example, when I watch a poker game and try to figure out whether a player should raise, call, or fold, I make use of my folk psychology. I look at other players, observe their bodily movements and interactions with one another, and try to read their faces in order to determine how good their hands are. And I also try to make predictions about what the other players will do next. But primarily, I want to know who is being deceptive and who is being honest with their bets. The end of this analysis isn't behavior, since I am not playing, but rather coming to some conclusion about the best description of the situation. Nonetheless,

the act of coming to that conclusion is a folk psychological act. Why? Because I am trying to determine something about others' behaviors, moods, honesty, and so forth.

Another thing our folk psychology allows us to do is to *coordinate* behaviors with others. Consider for example a driver who tries to coordinate his driving behavior with others on the road so as to avoid accidents, or consider simply walking down a crowded city street. Coordination problems like these involve multiple agents who are all trying to achieve some goal. A strategy we use for solving such problems, according to Adam Morton, is to recognize each other's goals, rather than focusing on the contents of one another's minds (Morton 2003). Given the situation and the goal, we can anticipate the appropriate behavior.

Our folk psychology also aids us in *detecting deception*, so we use it not just for cooperation but for *competition* as well. Competition and coordination, which are considered key to the evolution of sophisticated human cognition, help us survive in social worlds whether they are full of friends or full of enemies. We coordinate actions with our friends, and we check for deception in order to compete with those who may be adversaries. In deception, as in cooperation, analysis at the level of belief/desire attribution can lead to a perhaps unanalyzable degree of recursive complexity.

Perhaps one of the most important folk psychological practices (but least discussed in the philosophical literature) is the practice of *bonding* with other individuals. Bonding requires us to create and sustain emotionally intimate relationships with relatives and those who become our friends, and an essential part of these relationships involves attempts at *understanding*. It is our ability to understand that allows us to know when our intimates need to be comforted, want to play, or when they could use some distraction. All these folk psychological practices, and others, are important because they allow us to take part in society. For humans, such folk psychological practices may have been the most important in our development into a sophisticated technological species. Psychologists have suggested that the understanding we arrive at via our folk psychological practices may serve as a kind of glue that holds society together. Folk psychology qua the forming of emotional relationships and caring about others' behavior (and gossiping about it) may be the roots of human intelligence, culture, and all that goes along with it (see, e.g., Dunbar 1996; Greenspan and Shanker 2004; de Waal 1996).

The common folk psychological practices of predicting and explaining behavior are just the tip of the iceberg. When other acts, such as coordinating, cooperating, justifying, bonding, evaluating, etc., are considered, it seems clear that they too play an important role. And, if the function of folk psychology is to help us interact with other agents in our environment by doing things such as predicting, explaining, coordinating, judging, etc., then folk psychology cannot be merely the attribution of beliefs and desires, as the textbook version puts it. Attribution of beliefs and desires is not sufficient for any one of these practices, nor is it a plausible complete account of any one of these practices, including explanation.

What follows is that the main descriptive claim of the standard account, i.e., that we must be able to attribute beliefs and desires in order to predict behavior, is false. Humans don't need to appeal to mental states in order to predict and explain behavior, at least not in every case. The standard account doesn't begin to capture what humans do when they engage in their folk psychological practices, nor does it capture how children develop their folk psychological practices. It is widely accepted that children do not begin to attribute beliefs and desires that are different from their own until age 3½ to 4. However, before this time children engage in a host of behaviors that can only be taken as evidence for some understanding of psychological phenomena (see Greenspan and Shanker 2004 for a discussion). As early as 4 months infants begin to engage in emotional signaling with a caretaker by participating in bouts of back-and-forth interactive verbalizations or behaviors. When toddlers are still using only single-word utterances, from 9- to 18-months, they are developing a sensitivity to the moods of their caretakers, and will respond differently according to a person's mood. For example, if a child's mother is sad, the toddler may attempt to console her by giving her a favorite toy. If a child's father is angry, the toddler may hide. In the second year of life, children are already predicting people's behavior depending on the target's mood. They also demonstrate understanding about the causes of others' moods and knowledge about how to change them. For example, children as young as 1½ appear to initiate reconciliation with their parents after having a tantrum (Potegal and Davidson 1997). All these behaviors demonstrate social cognition skills, including prediction, and this is before the child has any significant language skills, before she passes the false belief task, and before she has a robust concept of belief. Nonetheless, in these cases there is some conception of psychological phenomena. The young children are reacting as agents to agents.

Again, given the social psychological and the developmental evidence, folk psychology cannot be the attribution of beliefs and desires to predict and explain behavior, unless one wants to both preclude children from having a folk psychology until age 4, despite their being highly social, and to accept that humans rarely use folk psychology to predict behavior. One might insist on this definition, and accept the consequences of it. While those who want to limit folk psychology to the prediction and explanation of behavior based on the attribution of propositional attitudes are free to do so, they owe us an account of why this subset of our understanding of psychological phenomenon is theoretically relevant, given the evidence from social and developmental psychology. Folk psychology understood narrowly doesn't exhaust the adult human's methods of predicting behavior. Since our predictions are generally accurate and since research suggests our predictions based on belief/desire attribution are not so accurate, there is reason to suspect that such attribution is not a particularly common method of predicting behavior. Nor does folk psychology understood narrowly capture the various methods adult humans use to explain behavior, such as explaining by using situational facts like historical precursors and enabling conditions (Andrews 2003; Malle 2004).[5] Given that folk psychology was supposed to correctly describe at least those two practices,

the standard account of folk psychology fails to deliver on its promise. It is for this reason that we ought to reject the standard definition, and rather begin to develop a wider conception of folk psychology. Our examination of animal folk psychology can help us determine what a wide conception of folk psychology might look like.

11.3. CRITTER PSYCHOLOGY

A critter psychology requires us to look not at how *humans* understand animal minds, but at how the *animals themselves* understand other minds (if they do at all). If animals do understand other minds, then there are bound to be differences between their understandings and our own. Given the criticisms of homogeneity and narrowness of function, we should modify our original question so that it is more specific. First, for a moment we can leave aside the question of the method used for our folk psychological practices, and ask whether animals other than humans engage in any folk psychological practices, such as predicting, explaining, coordinating, justifying, bonding, and evaluating. To have a folk psychology, one needs to engage in at least some of the folk psychological behaviors. Do animals engage in such behaviors? The answer to the question is easy: Sure, they do. Some of them.

From Darwin's anecdotes to Donald Griffin's accounts of animal behavior, and throughout the current literature, there is plenty of evidence that many different animal species engage in some of the behaviors associated with folk psychology (e.g., Bekoff and Byers 1998; Byrne and Whiten 1988; Cheney and Seyfarth 1990; Griffin 1992; de Waal 1982, 1996). Animals clearly *predict* the behavior of their conspecifics, predators, and prey. Without this ability, individuals would soon be dead, and species would soon be extinct, since predicting allows animals to engage in such essential behaviors as mating, eating, and avoiding being eaten. Anecdotes and formal studies have provided evidence of behaviors consistent with folk psychological practices such as competition, deception (at least tactical deception), justification, punishment, social bonding, and more. Let's look briefly at the evidence for one important folk psychological practice – coordinating behavior – in one species, the chimpanzee.

Chimpanzees live in stable social groups that allow the animals to develop long-term and dynamic relationships with other members of the group. Among the practices that the apes engage in are food sharing, exchanges of services such as grooming, deference to authority, and tactical alliance formation (often based on these other criteria). In addition to these behaviors, and sometimes as part of them, chimpanzees will coordinate behavior. One simple example can be seen in the dominance hierarchy of a social group. The dominant animal might have one or two animals he relies on for tactical support, and these animals might help the dominant fight off attacks by usurpers, or they might turn on the dominant if joined by a willing (and capable) assistant. These relationships involve coordination of behavior in order to achieve a common goal, be it the support or the downfall of the alpha (de Waal 1996).

An example of sophisticated behavior coordination can be seen in the chimpanzees of the Taï Forest of Cote d'Ivoire, who engage in a highly complex cooperative hunting strategy. These chimpanzees also follow meat-sharing rules whereby the amount of food an individual receives depends on the role(s) he performed during the hunt (Boesch 2002). Typically, there are four roles that the animals could take on in a hunt for monkeys: driver, chaser, ambusher, and captor. When the prey is spotted, the hunters take on one of these roles based on their location in relation to the monkey and their anticipation of the monkey's behavior. The hunters have to behave flexibly, for they will change roles as the situation dictates. Each of these roles is rather difficult to perform, and it can take a Taï chimpanzee 20 years to become a proficient hunter.

Given this sampling of the empirical evidence, let us now premise our deliberations with the claim that at least some animals engage in folk psychological practices such as predicting, coordinating, and bonding. If we identify folk psychology as the ability of intentional agents to engage in these behaviors, Argument C can be reformulated as follows:

Argument C′

1. Any intentional agent who engages in behaviors including predicting, explaining, coordinating, detecting deception, bonding, understanding, and justifying has some kind of folk psychology.
2. Some animals are intentional agents who predict behavior, coordinate behavior, and bond with other individuals.
3. From (1) and (2) it follows that some animals engage in folk psychological practices, and hence have some kind of folk psychology.

If folk psychology is defined as the ability to engage in behaviors that can be interpreted as folk psychological, then it would follow that many animal species have a folk psychology. But like our first argument, Argument C′ also seems a bit cheap, because the traditional question about folk psychology, for humans as well as for animals, was thought to involve fundamentally the attribution of mental states. This issue is completely avoided in a purely functional analysis of folk psychology. It seems we are in a quandary. The traditional question is to ask whether animals attribute beliefs and desires to predict behavior. But we saw that the attribution of beliefs and desires need not be necessary for humans to predict behavior, and that children engage in folk psychological behaviors before they have an understanding of belief. Thus, it seems that we cannot require that animals attribute beliefs in order to have a folk psychology.

To avoid this quandary we must determine whether there is something important about folk psychology that is captured by the standard definition, something that we should preserve in a new account. Engaging in some social behaviors and being an intentional agent are both necessary for one to have a folk psychology, but unless Argument C′ satisfies, they can't be sufficient. What is clearly missing from the definition of folk psychology in premise (1) is the criterion that the agent understands others as minded, intentional creatures. Having this sort of knowledge about other minds is part of the standard definition, and we can preserve that aspect

without insisting that the understanding of others as minded *requires* the concepts of belief and desire. There is some other understanding of others as intentional agents that children develop when they begin to engage in folk psychological practices. As children learn to respond differently to an angry mommy and a happy mommy, they see the caretaker as a minded agent, even though they don't yet have an understanding of belief. To see someone having a mood, feeling a pain or desiring a cookie is to see an intentional agent.

Rather than insisting that the ability to attribute both beliefs and desires is necessary for having a folk psychology, we can look at both the mental concepts demonstrated by adults in our various folk psychology practices, and the mental concepts demonstrated by children who are developing a more robust folk psychology.[6] Children younger than 3½ don't attribute beliefs, but they do have a folk psychology based on mentalistic concepts such as desire, seeing, intentionality, and so forth (Wellman 2002). They know that people seek out things that they desire, and avoid things that disgust them. And they know this as early as 18 months (Repacholi and Gopnik 1997). Children use their knowledge about other people's minds, moods, desires, etc. to make predictions about people's behavior. By the time we develop into adults, humans have additional mental concepts such as personality traits and beliefs.

In addition to the mental concepts that we see discussed in social psychology and developmental psychology, we can also examine whether animals have any other intentional attitudes such as seeing, feeling, liking, hating, etc.. And finally, we can examine qualitative states such as hunger, discomfort, softness, and so forth. Anyone who attributes personality traits, intentional attitudes, or qualitative states to other agents, and who engages in some folk psychological behaviors, seems to fulfill the basic intuitions that led to the construction of the standard account of folk psychology.

To ask whether animals understand any mentalistic concepts is much closer in spirit to the question first asked by Premack and Woodruff (1978) about whether an animal has a theory of mind, given that their study focused on the mental state of seeing. However, in asking whether the chimpanzee has a theory of mind we must be careful to avoid the mistakes made in the traditional approaches to answering this question. We should avoid looking for mentalistic concepts solely in acts of prediction, given that we know humans can make predictions without appeal to the target's mental states. And we need to shift focus away from beliefs and desires, without ignoring such concepts all together.

Before we attempt to answer the question about whether any animal has a mentalistic concept, a few words should be said about why psychologists think animals have *any* concepts. The major players in the debate over chimpanzee theory of mind agree that chimpanzees have some nonmentalistic concepts, such as 'same' and 'different.' The evidence for this conclusion comes from a number of standard research paradigms, such as the paradigm that requires subjects to judge whether a novel pair of objects is the same or different. If the objects are different from the objects the chimpanzee is trained on, the chimpanzee's ability to succeed at this task cannot be explained by the animal's association between stimulus and response.

Rather, the chimpanzee is able to generalize from a training set to novel objects, thus suggesting that the animal's response is made possible by abstract conceptualization. The chimpanzee's success in this and other experimental paradigms such as stimulus equivalence, sorting based on perceptual and functional features, and transitivity tasks is taken as sufficient evidence for a variety of chimpanzee concepts (see, e.g., Call 2001; Povinelli and Vonk 2004).

Our question, then, is whether animals have any mental concepts. This question is subtly different from Premack and Woodruff's original question about theory of mind. For Premack, and for those who followed, 'theory of mind' refers to the ability to "attribute states of mind to [others], and use these states to predict and explain the behavior of [others]" (Premack 1988, p. 160). Note that this definition makes the narrowness assumption: To have a theory of mind is to predict and explain behavior by attributing mental states. There is very little in the way of consensus about whether any nonhuman animals do have a theory of mind. Because the majority of the research on this issue is on the great apes, primarily chimpanzees (the subject of Premack's original question), I will continue to focus on that species in the discussion that follows.

Seeing is thought to be a good place to start looking for mentalistic concepts other than belief and desire, given ethological evidence that chimpanzees monitor gaze and modify their behavior when they are visible to others. In 1978 the ethologist F.X. Plooij reported that wild chimpanzees glance at other animals' faces, presumably to check whether others can see them. This behavior develops over time; infant chimpanzees, like infant children, are not sensitive to the gaze of others, and will gesture to their mothers with requests for grooming without first looking to see if their mother is attending to them. But around 10½ months we see a change in the behavior of young chimpanzees, who begin to attend to their mothers' gaze before making a request.

There are other naturalistic behaviors that are suggestive of chimpanzee sensitivity to others' ability to see. For example, low-ranking primates tend to mate and feed outside of the sight of dominants (Whiten and Byrne 1988), and will suppress vocalizations that accompany forbidden behavior (such as sex between a subordinate male and a female preferred by the dominant) (Goodall 1986). In some cases, when a third party observes these secretive acts they will alert the dominant, who then moves in to interrupt the behavior. Chimpanzees also seem to recognize that letting others see their fear behavior is to be avoided. In one case, a chimpanzee began fear-grinning in response to threatening vocalizations by another chimpanzee. The rival couldn't see the fear-grinner, who was facing away from him. Before turning around to face his rival, the fear-grinner used his hand to pull his lips down over his teeth to stop the facial expression. It took him three tries, but after he succeeded in wiping the fear-grin off his face, he turned to confront his rival (de Waal 1996). Having a concept of seeing seems to help chimpanzees play politics, and generally engage in desirable behaviors while avoiding negative consequences.

Laboratory studies have been of mixed success in determining whether or not chimpanzees have a concept of seeing. Some early studies, such as those done by David Premack, suggested that chimpanzees do understand seeing. Another set of experiments led others to conclude that chimpanzees do not understand seeing (Povinelli and Eddy 1996). Most recently, a new set of studies suggests that chimpanzees understand both seeing and intentionality (Hare et al. 2000; Hare et al. 2001).

In Hare et al.'s (2001) experimental set-up, a subordinate and a dominant chimpanzee are released in a room baited with food. Normally, if both animals can see the food, or see one another witness the baiting, a subordinate animal will avoid the food and allow the dominant access, since dominants are known to punish subordinates for eating food without permission. However, in these experiments, when the food is occluded from the dominant's view, the subordinate will approach it. Only if the dominant can see the food will the subordinate avoid it. The animals are across the room from one another, so the subordinate has to consider the visual perspective of the dominant in order to judge correctly whether he can see the food or not. Because it seems that the subordinate is able to make different judgments about whether to seek out the food based only on whether it is visible to the dominant, this study is thought to indicate that the apes have a concept of seeing.

In subsequent discussion of these results, one criticism is that as a predictive paradigm, the subordinate could be using a nonmentalistic strategy to predict the behavior of the dominant. It has been argued that performance on this task can be explained by the fact that chimpanzees (like humans) probably form abstract representations/categories and general rules about the behavior of others that are used to predict others' behavior and to modify their own behavior accordingly (Povinelli and Vonk 2004). Recall that induction over past behavior is one method that humans use to predict behavior, and Povinelli and Vonk suggest that the chimpanzees could be using just this technique.

But, the authors of the study point out that this interpretation won't work, because not all experimental conditions involve reading behavior. To test whether the subordinate could be using a rule such as, "Avoid food when the dominant is bodily orientated toward it, or when his head is turned toward it, otherwise approach food," the experimenters have a condition in which the dominant isn't visible until after the subordinate has already started moving. Because the dominant is hidden behind an opaque door, and is released only after the subordinate makes his decision whether to head toward the food or not, the subordinate has no behavior to read. In this condition, the subordinate continues with the same behaviors as before: When the food is obscured from the dominant's point of view at his doorway, the subordinate approaches the food. When the food is visible from the dominant's doorway, the subordinate refrains from approaching the food. Thus, it seems that there could be no decision-making purely on the basis of behavioral abstractions. In addition, when the food is obscured from the subordinate, but visible to the dominant, and the subordinate is given the opportunity to observe the dominant's behavior before being released into the room, the subordinate is not able to determine

where the food has been placed. So it seems unlikely that the chimpanzees are able to make judgments about whether or not the dominant would seek out food simply from the way he moves his body (Tomasello et al. 2003). As another control, when both the dominant and subordinate are able to watch the baiting of the room, the subordinate will avoid all food, even when it is placed behind a barrier and no longer visible to the dominant. Given the results of these conditions, it is highly unlikely that the subordinate can make a generalization based on the dominant's body positioning.

If the subordinate isn't able to make his decision whether or not to seek out the food based on behavioral abstractions, how does he do it? Perhaps the subordinate, knowing that the dominant is going to be released, uses inductive reasoning to conclude that the dominant will move his body in certain ways if he is released. Knowing that the dominant will move his body in a certain way, perhaps the subordinate can predict that the dominant will then seek out the food. There is one problem with this nonmentalistic account of the subordinate's reasoning, namely that the question is simply pushed back a step. Given this response, we still need to know what makes the subordinate anticipate that the dominant will move his body in certain ways. The answer to that question may be where the mentalistic reasoning lies.

Suppose the subordinate doesn't understand seeing, or the visibility of objects, and he fails the first couple of trials. Then, given his past experience, he knows only that when the door opens, sometimes the dominant first moves his body toward the food, and then when released heads toward the food to eat it. But sometimes the dominant doesn't move his body toward the food, and he doesn't seek it out when released. At this point, the subordinate is faced with a puzzle he must explain.[7] He needs to find out what the relevant difference is between the two conditions. He cannot make a straightforward prediction based on inductive generalizations over past behavior, trait attributions, or stereotype. What needs to be explained is why the dominant heads toward the food in some cases, but not in other cases. To answer that question, the subordinate must look at the context of the situation. He must look at the food, the barriers, and find some rule that connects the set-up of the situation in all the cases the dominant seeks out the food, and all the cases where he doesn't. The subordinate doesn't have a simple rule about how behavior differs based on the location of the food. What he needs to do is make an abstract generalization about the property of the food in relation to the barriers in the different conditions. And it is the abstract property of invisibility that leads the subordinate to anticipate that the dominant will not approach it.

With the ability to respond to the abstract concepts of visibility/invisibility, the subordinate can categorize the food item as one that the dominant will approach versus one that the dominant will not approach. No nonmentalistic concept will do. The only thing in common with all the cases where the subordinate avoids the food is its visibility to the dominant, and so the only appropriate abstraction is the visibility of the food. Visibility is a mentalistic concept; in order to categorize

the food as visible or not visible by the dominant, the subordinate must understand something about seeing.

Despite this conclusion, there is reason to think that the chimpanzee's concept of seeing is different from the adult human concept. The human concept of seeing includes the idea that seeing leads to believing. This point is emphasized in the child theory of mind literature. According to the human concept of seeing, someone who sees something happen will then be in a doxastic state regarding that event, whereas one who does not witness the event will not share that doxastic state, *ceterus paribus*. The child who passes the false belief task is described as using the following pattern of reasoning: he knows that Sally didn't see Ann move the marble, and therefore he knows that Sally doesn't believe that the marble is in the cupboard. Given that Sally will act on her beliefs, the child predicts that Sally will seek out her marble in the box, where she believes it is. Note that though passing this task is taken to be evidence of having the concept of belief, the reason we understand it this way is because our concept of belief is related to our concept of seeing. What the child who passes the false belief task knows is that Sally didn't *see* the marble being moved, and from that we presume the child knows that Sally doesn't *believe* that the marble was moved. In adult human reasoning, seeing and believing are two categories that are closely tied. In children, despite the standard interpretation of the false belief task, they may not be.

Our concept of seeing has behaviors associated with it, in addition to having connections with other concepts. So, while humans and chimpanzees may take seeing as referring to roughly the same sorts of behaviors (such as reaching for a desired seen object, running from a threatening seen object, and so forth), there is little current evidence that chimpanzees and humans make the same connections between different concepts; there is little experimental evidence that the chimpanzee concept of seeing is related to a concept of believing. Since in humans the category of behaviors we call seeing is associated with the doxastic category of believing, in order to say that a chimpanzee has our concept of seeing we would need to determine whether that concept is connected to a chimpanzee concept of believing. Given these considerations, what should we conclude about the chimpanzee's ability to make judgments about what others can and cannot see? Despite the concern that it might be too quick to describe the chimpanzee's concept as identical to the adult human concept of seeing, it does seem clear that the chimpanzee uses an abstract mentalistic concept in order to solve puzzles and make complex predictions. Further, it seems that the chimpanzee concept of seeing is not unlike the concept of seeing a child has before her fourth birthday.

If we can take the above evidence as sufficient to claim that chimpanzees have some understanding of seeing, then we can incorporate that fact into our argument about animal folk psychology. The question we began this section with was whether animals, like humans, think that their conspecifics are mentalistic agents. Since seeing is a mentalistic concept, and it seems that chimpanzees attribute something like seeing to other chimpanzees, and use the concept in predicting, coordinating, and competing, we could conclude that the chimpanzee does think of

other chimpanzees as intentional agents. So we can reformulate our argument for animal folk psychology in terms of chimpanzee folk psychology as follows:

Argument C″

1. Any intentional agent who engages in folk psychological practices and understands others as minded agents has some kind of a folk psychology.
2. Chimpanzees are intentional agents who engage in folk psychological practices and have some understanding of others as minded agents.
3. From (1) and (2) it follows that chimpanzees have some kind of folk psychology.

On the question of whether chimpanzees might have something analogous to human folk psychology, with the understanding that human folk psychology involves two elements – practices of social interaction, and abstract mentalistic concepts – there now seems to be reason to accept that chimpanzees do have some kind of folk psychology. For one, the definition of folk psychology used above captures both the behavioral and the conceptual aspects that are associated with human folk psychology. However, the argument is limited insofar as it rests on controversial evidence (the seeing studies), a limited mentalistic concept (only seeing), and a limited class of animals (chimpanzees). And, it is important to note that this argument doesn't entail that chimpanzees have the same folk psychology as humans. Given the above discussion about the chimpanzee concept of seeing it seems likely that the chimpanzee would not have the same folk psychology as adult humans. If this is the case, then there is no one monolithic folk psychology to be had by all, contrary to what some might think. This examination of nonhuman folk psychology should lead us to rethink some assumptions about human folk psychology, such as the universality of our social cognition. If there are differences in folk psychology across species, there may be folk psychological dialects across human cultures too. In order to determine whether there is only one conceptual scheme when it comes to the minds of other people, additional comparative research in social psychology is required.

To broaden the conclusion of Argument C″, research must be done on other mentalistic concepts. Techniques for finding mental states concepts among animals might follow two approaches. First, if ontogeny recapitulates phylogeny, we should look toward the development of mental state concepts in children. Belief concepts are acquired relatively late by children. They begin to develop around age 3½, but children don't have the full adult concept of belief immediately after they pass the false belief task. The understanding of belief continues to develop over the next several years as children begin to develop an understanding of the many properties of belief, such as the opacity of belief ascriptions (Apperly and Robinson 1998, 2003). Plenty of other mental state concepts develop much earlier, and these are the mental states we should start looking for in nonhuman animals. Those concepts that children develop first, emotional concepts such as 'pleasure,' 'comfort,' and 'pain,' would be a good place to start.

Second, given that our analysis of human folk psychology has undermined the traditional emphasis on prediction, nonpredictive experimental paradigms should be explored, as should evidence from ethological observations. If mental states

are used by humans to explain intentional behavior, and if humans tend to seek explanations for behaviors they find anomalous, one approach to examining animals' understanding of mental concepts would be to expose them to anomalous behavior that has a simple mentalistic explanation behind it. If the animals seek to find an explanation, that will serve as some additional evidence that they have the mental state concept.

11.4. CONCLUSION

The main objection to my argument for a variety of folk psychology in chimpanzees will come from those who insist that folk psychology requires a full-blown theory of mind. According to this objection, folk psychology is a robust conceptual scheme that must involve many mentalistic concepts including belief and can be had only by those who explain behavior. Since there is no evidence that nonhuman animals attribute beliefs, or that they have a theory of mind, they shouldn't be said to have a folk psychology.

These kinds of criteria for having a folk psychology are too strong, given the general account as the commonsense understanding of psychological phenomenon. First, though a theory of mind is traditionally seen as the attribution of beliefs and desires, we have seen that human social cognition does not place emphasis on belief/desire attribution. The objection falls prey to an oversimplified view of the mechanisms underlying human folk psychological abilities. There is growing recognition that there isn't any one specific mechanism that underlies all our folk psychological practices, but that different practices will tend to rely on different mechanisms, as will different instances of the same kind of practice. While humans do attribute beliefs and desires as part of some of our folk psychological practices, especially explaining behavior, we have seen that humans don't need belief/desire attribution in order to engage in other practices, e.g., in many cases of predicting behavior.

Some techniques humans use to predict behavior, such as generalization from past behavior, or using trait attributions, are almost certainly used by chimpanzees as well. The well-defined social relations of chimpanzee groups depend not only on knowing what animals have done in the past, but whose party they belong to. It is also useful to know whether an animal tends to share food (generous) or hoards it to himself (selfish). Such individual differences are common in chimpanzee societies, and knowing about them is useful for group members who are interested in maximizing their position in the group. We know from ethological observations that chimpanzees treat animals we might consider selfish differently from those we would describe as generous, for example (Watts 2002).

The definition of folk psychology used in Argument C″ emphasizes not only the fact that there are a variety of folk psychological practices and a variety of mental concepts, but also that having a folk psychology comes in degrees. Any creature that fulfills the two requirements of the definition of folk psychology from premise (1) of Argument C′, engaging in folk psychological practices and having folk psychological concepts, should count as having some kind of folk psychology.

It need not be as robust as ours, nor as complex, but anything that uses mentalistic concepts to engage in social interaction counts. To claim otherwise is to reject the definition.

And we shouldn't reject the definition. Children younger than 3½ don't have a full-blown theory of mind, but they have a folk psychology based on mentalistic concepts such as desire, seeing, intentionality, and so forth. They know that people seek out things that they desire and avoid things they hate. They know things about how people's moods affect their behaviors, and they use this information to predict people's behavior. Again, it seems clear from our definition that knowledge of mentalistic properties such as desire, disgust, pleasure, and so forth allows children to interact in the social domain, and that is sufficient to show that they land on the folk psychological continuum. I hope to have convinced you that some of our nonhuman relatives land on this continuum too.

NOTES

[1] For Dennett (1991, p. 29), the intentional stance is an easy and reliable method for predicting behavior, and more. The predictive power of folk psychology makes possible all our interpersonal projects and relations; without it, "we would be baffling ciphers to each other and to ourselves".

[2] While Goldman (1995, p. 88) does think that people can make predictions of intentional agents by means of "generalizations or other inductively formed representations (schemas, scripts, and so forth)," these methods are derivative of simulation, and can only be used by a "mature cognizer" who has already engaged in many acts of mental simulation.

[3] This story is consistent with the general views associated with both the simulation theory and the theory theory accounts of the form of folk psychology. (However, Robert Gordon's rather radical simulation, which involves attending only to the environment rather than the agent, would be excluded by this definition.) It is clear that theory theory implies that we predict and explain behavior by attributing mental states, but it may be less apparent that simulation theory makes use of mental state attribution. Nonetheless, according to most views of simulation the attribution of mental states is necessary to get a simulation started, according to simulation theorists such as Alvin Goldman. When beginning a mental simulation, he suggests we assume that the agent is like us, that we share relevant psychological features. This assumption makes us accept explanations that seem natural to us, and reject those that are less natural (Goldman 1995). Goldman even suggests that the simulator should assume shared basic likings and desires, unless there is reason to think otherwise. In order to make these assumptions, we must attribute propositional attitudes (e.g., Bill wants to stay alive, Bill doesn't want to be injured, Bill believes that being injured causes pain, etc.).

[4] The primary problem with trait attribution as a method of prediction is that it seems to function only in limited situations, because people's cross-situational consistency is not nearly as great as we expect. So, while your colleague might be generous with her money, she may be stingy with her time, or while she may be generous around the office, she may not be generous with her friends and family. See Ross and Nisbett (1991) for a review of this literature.

[5] For example, I can explain why Henry Rollins did a Gap advertisement by saying that after living in poverty for so long he sold out. I might not know what he told himself to rationalize the action, and he might not agree that his poverty played a causal role. He probably would disagree that he sold out. Nonetheless, it is natural to give such an explanation, even though it does not attempt to describe the agent's own reason explanation. Not all psychological explanations are Davidsonian rationalizations.

[6] I should note that in this section I am using the term "concept" very loosely, as whatever it is that allows one to abstract, and hence to generalize and categorize. This is in keeping with the standard psychological notion of concept, which operationalizes the term as the ability to categorize and make discriminations. However, I am uneasy about some of the more specific commitments some psychologists

have about the nature of concepts. For example, Josep Call argues that the chimpanzees' concepts are 'intervening variables' that are used to construct and apply general rules that can be applied to novel situations. I am concerned that this view is a misrepresentation of both the human's and the chimpanzee's use of concepts. However, the current argument remains neutral on this issue.

[7] I don't mean to suggest that the chimpanzees in this study developed a concept of seeing in the course of the experiment. The point I am making here is that puzzling behaviors can lead an individual to discover mental states more generally, and it is this developmental claim that I mean to emphasize.

REFERENCES

Andrews K (2003) Knowing mental states: the asymmetry of psychological prediction and explanation. In: Smith Q, Jokic A (eds) Consciousness: new philosophical perspectives. Oxford University Press, Oxford.

Apperly IA, Robinson EJ (1998) Children's mental representation of referential relations. Cognition 67:287–309.

Apperly IA, Robinson EJ (2003) When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year old children's reasoning about beliefs and belief reports. J Exp Child Psychol 85:297–311.

Bekoff M, Byers J (1998) Animal play: evolutionary, comparative and ecological perspectives. Cambridge University Press, Cambridge.

Boesch C (2002) Cooperative hunting roles among Taï chimpanzees. Human Nat 13:27–46.

Byrne R, Whiten A (eds) (1988) Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans. Oxford University Press, New York.

Call J (2001) Chimpanzee social cognition. Trends Cogn Sci 5:388–393.

Cheney DL, Seyfarth RM (1990) How monkeys see the world: inside the mind of another species. The University of Chicago Press, Chicago.

Churchland P (1981) Elimiative materialism and the propositional attitudes. J Philos 78:67–90.

Dennett D (1987) The intentional stance. MIT Press, Cambridge, MA.

Dennett D (1991) Real patterns. J Philos 88:27–51.

Dunbar R (1996) Grooming, gossip, and the evolution of language. Harvard Univeristy Press, Cambridge, MA.

Fodor J (1989) Psychosemantics: the problem of meaning in the philosophy of mind. MIT Press, Cambridge, MA.

Goldman A (1995) Interpretation psychologized. In: Davies M, Stone T (eds) Folk psychology. Blackwell, Oxford.

Goodall J (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge, MA.

Greenspan SI, Shanker SG (2004) The first idea: how symbols, language, and intelligence evolved from our primate ancestors to modern humans. Da Capo Press, Cambridge, MA.

Griffin DR (1992) Animal minds. University of Chicago Press, Chicago.

Hare B, Call J, Agnetta B, Tomasello M (2000) Chimpanzees know what conspecifics do and do not see. Animal Behav 59:771–785.

Hare B, Call J, Tomasello M (2001) Do chimpanzees know what conspecifics know? Animal Behav 61:139–151.

Kreuger J (1998) On the perception of social consensus. Adv Exp Social Psychol 30:163–240.

Kunda Z (2002) Social cognition: making sense of people. MIT Press, Cambridge, MA.

Locksley A, Borgida E, Brekke N, Hepburn C (1980) Sex stereotypes and social judgment. J Personal Social Psychol 39:821–831.

Locksley A, Hepburn C, Ortiz V (1982) Social stereotypes and judgments of individuals. J Exp Social Psychol 18:23–42.

Malle BF (2004) How the mind explains behavior: folk explanations, meaning and social interaction. MIT Press, Cambridge, MA.

Marks G, Miller N (1987) Ten years of research on the false-consensus effect: an empirical and theoretical review. Psychol Bull 102:72–90.
Morton A (2003) The importance of being understood: folk psychology as ethics. Routledge, London.
Mullen B, Atkins JL, Champion DS, Edwards C, Hardy D, Story JE, Vanderlok M (1985) The false consensus effect: a meta-analysis of 115 hypothesis tests. J Exp Social Psychol 21:262–283.
Potegal M, Davidson RJ (1997) Young children's post tantrum affiliation with their parents. Aggress Behav 23:329–341.
Povinelli DJ, Eddy TJ (1996) What young chimpanzees know about seeing. Monogr Soc Res Child Dev 61.
Povinelli DJ, Vonk J (2004) We don't need a microscope to explore the chimpanzee mind. Mind Lang 19:1–28.
Premack D (1988) 'Does the chimpanzee have a theory of mind?' revisited. In: Byrne R, Whiten A (eds) Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans. Oxford University Press, Oxford.
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 1:515–526.
Repacholi BM, Gopnik A (1997) Early reasoning about desires: evidence from 14- and 18-month-olds. Dev Psychol 33:12–21.
Ross L, Nisbett RE (1991) The person and the situation: perspectives of social psychology. Temple University Press.
Tomasello M, Call J, Hare B (2003) Chimpanzees versus humans: it's not that simple. Trends Cogn Sci 7:239–240.
Von Eckardt B (1994) Folk psychology. In: Guttenplan S (ed) A companion to the philosophy of mind. Blackwell, Cambridge, MA, pp 300–307.
Waal Fde (1982) Chimpanze politics: power and sex among apes. Jonathan Cape, London.
Waal Fde (1996) Good natured: the origins of right and wrong in humans and other animals. Harvard University Press, Cambridge, MA.
Watts DP (2002) Reciprocity and interchange in the social relationships of wild male chimpanzees. Behav 139:343–370.
Wellman HM (2002) Understanding the psychological world: developing a theory of mind. In: Goswami U (ed) Blackwell handbook of childhood cognitive development. Blackwell, Oxford.
Whiten A, Byrne RW (1988) The Machiavellian intellect hypotheses. In: Byrne RW, Whiten A (eds) Machiavellian intelligence. Oxford University Press, Oxford.
Wilson TD, LaFleur SJ (1995) Knowing what you'll do: effects of analyzing reasons on self-prediction. J Personal Social Psychol 68:21–35.

ADAM MORTON

12. FOLK PSYCHOLOGY DOES NOT EXIST

> Guy was realising more and more that it wasn't just hard to put yourself in another's mind, but nearly impossible, although that was supposedly part of the acting profession. The truth was that you absorbed traits rather than mentality. In plays and scripts you always had tracks of cause and effect. But in life if you were dealing with people who didn't come from your own patch you weren't going to get it right. The answers came haphazardly, from the spinning wheel of a roulette table.
>
> Tibor Fischer 'Listed for trial'

12.1. INTRODUCTION

There have been many disputes in philosophy and psychology in the past 25 years over the nature of something that is variously called "folk psychology", "theory of mind", "mindreading" and other things. (Those names are not in fact real synonyms; each carries a different load of presuppositions. For a brief history of the topic, and finer distinctions within it, see Morton forthcoming-*a*.) There can be a right answer to such questions only if we are dealing with something that might have a nature. If not, the questions we ask may be like "what is the true nature of luck?" or "let's find the essential properties of the constellations". In this paper I shall explore the possibility that we apply the "folk psychology" label to too varied a bundle of capacities and phenomena for there to be a single tidy account of it. Of course there still might be tidy accounts of particular capacities in the bundle, but we should then be very careful how we label them. They wouldn't be anything like general accounts of how people understand people.

My argument is not meant to be conclusive. My conclusion will be that the non-existence of folk psychology, as a single unitary capacity, is a possibility that we ought to take seriously. We are weighing conjectures here – as we usually are in philosophy, even when it presents itself as delivering the results of inescapable argument – and this conjecture is part of a larger and more conjectural thought, which I shall mention to give the discussion some perspective, and then not return to. Folk psychology is supposed to be the means by which people in ordinary life understand the minds of other people. "Mind" covers a lot of ground: motivation, belief, consciousness, emotion, character and more. How much unity does this list have, besides a vague causal link to the nervous system? Are questions about the nature of "the mental" and its relation to physical reality well-formed? I am not convinced, either way. It seems to me entirely possible that when future cultures try to read our books they will need long glosses on the scattered variety of contexts in

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 211–221.

which we talk of mind. (Just as when we philosophers go into popular bookstores and see a section marked "mind and spirit" we shudder and ask for the distinction to be explained to us. Many non-academics use "mind" to refer to a cluster based on intellect and character and spirit to refer to a cluster based on consciousness and emotion. Perhaps, just perhaps, that is a better way to do it.)

The structure of my argument is as follows. I first present a picture of the activities that folk psychology is supposed to perform that makes it clear that they can be performed by a very loosely connected bundle of abilities. I then discuss the role the folk psychological vocabulary can play in holding such a loose bundle together. I then connect the discussion with imagination-based or simulationist accounts of folk psychology. And then to end the chapter I extract a rhetorically definite position from the preceding pros and cons.

12.2. THE VARIETY OF FUNCTIONS

Folk psychology is supposed to apply in everyday analogues of the situations in which scientific psychological theories might be applied. (There's an optimism about psychology here, as well as about folk psychology.) The idea is that we have a need to predict and explain what other people do, and that in order to do this we have to attribute beliefs, desires, emotions and other states to them. So the picture, at any rate on the dominant "theory theory" account, is that we have a practical need to know what to do with regard to someone else and we meet this by gathering information about her, using this information to attribute states of mind to her, combining these attributions with general beliefs about how people operate to predict what the person will or may do and then using these as inputs to one's own decisions. (Things look somewhat different from the rival simulationist camp. I'll return to this.) The crucial thing is that we use a single capacity to get from information about people to our own social decisions. That is the central dubious assumption. Consider some of the many possible combinations of practical situation and information about another person.

– You are lost in a strange city and you want to know who it is safe to ask for directions.
– You are attracted to someone and you want to know how to get to know them.
– You have stolen something and you want to know if the shop assistant has seen you take it.
– Your plan to meet someone at a particular place and time has failed, and you want to know what to do to find them.
– You want to get on with your new boss and you need to find out whether to act obsequiously, critically, or with initiative.
– Your friend's partner has died but your friend is carrying on in a relatively normal way; you want some sense of what kind of a time it is for her.
– You are considering spending your savings on a very expensive house, and you want to know whether in 5 years' time there will be many other people willing to spend a correspondingly large sum if you decide to sell.

Note that in several of these the situation is described in folk-psychological terms, or more neutrally in terms of social situations and attitudes. With a certain amount of effort we could rephrase the descriptions in terms just of the evidence provided and the acts you are considering, though losing some of the intuitive content. The fact is that when we try to describe our involvement in social life we have no real choice but to use the folk-psychological vocabulary. (That's an argument *for* the reality of folk psychology: but see the Sect. 12.3.)

These situations are very varied. We can describe the way we manage them in terms of a somewhat less varied array of social capacities. Central among them are:

– *Self-preservation*: the capacity to know when people are dangerous or cooperative.
– *Solving coordination problems*: the capacity to choose the outcome to a social situation that most other people in it will also find.
– *Emotional contagion*: the capacity to pick up other people's moods when in their presence (see Goldie 1999.)
– *Reflective grasp of rationality*: the capacity to calculate from an explicit statement of other people's aims and information, the choice that it would be most rational for them to make.

The important fact about these capacities, and many of the others that are recruited to get us through situations like those listed above, is that they also have applications that are not natural candidates for folk psychology. Thus the self-preservation capacity can be served by processes that also tell you about fierce dogs, alligators and landslides. Skill with coordination problems is part of a general capacity to handle strategic choice, which applies for example when one is thinking through what outcomes it would be rational for a group of people to settle on, in terms of the facts of the case and what is in their interests, quite independently of their beliefs and desires. (It also can embody routines for giving approximate solutions to quite complex social situations, of which we acquire a large number, see Camerer 2003.) Emotional contagion can occur without attributing any state to anyone or forming any conclusions. One may not even know whose emotions one is picking up. And a reflective grasp of rationality is obviously something whose main application is in making one's own individual decisions, not second-guessing those of others. Moreover when we do apply it to predicting others we do so in very nuanced and constrained ways, knowing as we do how far anyone is from acting fully rationally on more than the rarest occasions. (I have elaborated the point about coordination problems in Chapter 1 of Morton (2003) and on the point about rationality in Chapter 2 of that book.)

One remarkable feature of several of the capacities we use to find our way through our relations with other people is that they result in something other than a belief: they are not directly linked to attribution or prediction. A skill in coordinating with others can manifest itself just in doing the appropriate thing, as when one is dancing with someone or playing a team sport. A capacity for sharing others' emotions can manifest itself in simply feeling something related to what they feel, with no

straightforward connection to any actions one performs as a result. The picture that emerges has two main features.

First, the beliefs we form about other people's minds and future actions result from a variety of capacities for social life and rational action that we all possess, in varying degrees. They are not the sole output of any of these processes, and in fact many of these processes often do not result in beliefs at all.

Second, the full human capacity to anticipate the actions of others and to attribute states of mind to them comes from combining these component skills. However many situations that call for anticipation or sympathy with others can be managed pretty well by using them singly.

Some hard empirical questions immediately arise. Do all people in a given culture manage their attitudes to others by use of the same component capacities? Do people in all cultures use the same bundle of capacities? If the answer to these questions were yes, then folk psychology would have a sort of a derivative real existence, as the thing you get when you combine X, Y, Z and W. My suspicion is that the answer is No. In particular, I suspect that some people and some cultures make very little use of explicit thoughts about rationality, and some people and some cultures make a lot more use of our capacities simply to do the appropriate or cooperative thing, without producing reasons for it. One factor that makes the differences between cultures hard to focus on is the fact that since people do not like others to be unpredictable, and sometimes react with extreme hostility to actions they find bizarre, we learn to act in ways that fit the capacities that those around have for reacting to us. Some of the limits of a local bundle of capacities will pass unnoticed because the behaviour that would reveal them has been discouraged. (One of the few people to have appreciated this point is Martin Kusch, see Kusch 1999.)

12.3. THE APPEARANCE OF UNITY

Separate though these capacities may be, when they are used in a deliberate and reflective way to mediate one person's dealings with another they are often connected and used in combination. The main tie between them is the vocabulary of mind, our talk of beliefs, desires, emotions, memory, reasoning and related concepts. We are good at making inferences between attributions in different parts of this vocabulary, extensive and varied as it is. "She thinks you stole her cat so of course she is angry at you", "he remembers when you were an arrogant young graduate student so he is somewhat defensive in your presence". And in particular, we can use very little of it without explicitly or implicitly introducing words for belief and desire. This is the most impressive and intuitive reason for believing in the solidity of folk psychology: the coherence and cohesion of the folk psychological vocabulary. It provides us with a way of organizing our thoughts and attitudes about ourselves and others that moves us beyond simple reactions and anticipations to structured thinking about personality and motive. No doubt without it organized human life could not exist.

I do not want to deny this. Indeed I think that the folk psychological vocabulary has a greater richness and a more complicated structure than many philosophers and psychologists assume. (They often speak as if it consisted of some assumptions about the relations between beliefs, desires and actions, plus a few optional extras.) And I think that one task of philosophers of mind is to explore and even improve the vocabulary, so that we can use it more easily and apply it to a larger range of cases. (In fact, philosophers have long been doing this, indirectly, in honing the language of rationality and our vocabulary for the emotions. Think of the emotions in the family containing guilt, remorse, shame and embarrassment, for which we do not have names or standard contrasts to differentiate them from the others. Some of them could very helpfully be introduced into our standard vocabulary. Others, depending on our social purposes, are best left un-named.) But without denying any of this we can also note a number of ways in which the application of the folk psychological vocabulary is not as central and basic to our understanding of others as it might seem.

When we describe verbally a person's state of mind, think in terms of this description, and come to some conclusion about the person, our thinking is usually constrained by factors that we cannot articulate in the standard vocabulary, but without which the thinking would be impossible. You want to know how someone is likely to travel downtown (perhaps you want to be sure not to be travelling with the person, as there is a conversation you don't want to have). You think "she likes scenery, and the view from the train is a lot nicer than the view from the bus, so most likely that is how she will travel". Without noticing, you have ruled out cycling, hitchhiking, asking a neighbour for a ride, running and other ways she might get downtown. This may be because your intuitive grasp of her personality rules these things out. Or it may be because your grasp of the social situation in the context of which she was travelling made arrival in any of these ways a bad move. Or it may have been something else. The pattern is general: *articulate folk psychological description operates courtesy of a background of possibility-eliminating factors, of potentially unlimited variety*. (In this connection see Bermúdez 2004.)

We can sometimes combine descriptions of people's motives with knowledge of how people generally behave to predict what those particular people will do. You take someone to be stingy and risk-averse, and so you predict that he will not spend the extra couple of thousand getting his house repainted, which might result in his selling it for considerably more and might also have no effect on the sale. But we do not do as much prediction as it may appear. After someone has done something we often come up with explanations of their action (which have the same general form as predictions, but which we would not have be so rash as to produce as predictions beforehand). One reason for the asymmetry between prediction and explanation (which is well known in the philosophy of science) is that after the fact we have evidence for additional relevant facts about the person, namely that she has acted as she did. Often this extra information triggers constraints on the possibilities, as described in the previous paragraph, that cannot themselves be put into folk psychological terms. One result is that the explanation is then contrastive:

It explains not why the person did the act absolutely, but why the person did this act rather than one of a limited range of alternatives. In the face of this, it is very hard to tell real insight-giving explanation from pseudo-explanation faked up to fit the facts as we know them, sometimes a verbal shell for the real sub-conceptual grasp of personality that gives us our real hold on what people are like and what they are likely to do. As a result, *much of our calculation of the transition from motives to actions results not in absolute predictions but in contrastive explanations, where the range of the contrast is set from outside folk psychological thinking*. (See Morton 2003, chap. 4, and Morton 1996.)

A third factor is the hidden ambiguity of "belief" and "desire". Philosophers often write as if we had clear concepts of two relations between a person a and a proposition p "a believes p" and "a desires that p". In real spoken English we use a great variety of words: thinks, suspects, is of the opinion that,…;wants, longs for, would like, has a yen for, … And we use contrasts between these words to indicate different kinds of belief and desire. Compare for example your desire to get home without getting soaked on a wet night when your car has broken down to your desire that a ticket you have in a five million dollar lottery prove to be the winner. There is a sense in which the latter desire is stronger, since you would walk 10 miles in the rain for five million dollars, and a sense in which the former is, in that you react with dismay to the prospect of not getting home promptly and dry and just shrug your shoulders at the possibility that your ticket will not win. Similarly, you may "believe" that your position on some philosophical position is right, in that you defend it with energy and ingenuity and, particularly, assert it with conviction. But you recognize that the arguments against it have force, and you are rather less confident of it than you are about many things, for example, that there is no life on Mars, that you would not claim to believe even though you find them fairly likely. (A creature with vastly greater cognitive powers than human beings might well not have any beliefs: It would give to each proposition it considered a degree of belief, and then act accordingly, without ever making a slice between the ones it said Yes to and all the rest.) When we ascribe beliefs and desires and use them to explain and predict what people are doing the ascriptions are incomplete: We rely on conversational context and our knowledge of the particular person involved, and no doubt other factors, to fill in the full content of the ascription.

This view of belief and desire ascriptions is controversial. (I have given a more thorough defence in Chapter 4 of Morton 2003.) Something very similar is quite obviously true of other parts of the folk psychological vocabulary, though, in particular our words for emotions and attitudes. Consider the family of regret, remorse, shame, guilt and embarrassment. There are differences between all of these, but we are usually not very careful about which one we use. We say "regret" when "remorse" would be more precise, or "shame" when "embarrassment" would apply better. There are many emotions intermediate between these terms. (Not all languages have terms for all of these, and the words for retrospective negative emotions in different languages rarely translate very exactly.) So when one person says of someone that he, for example, feels guilty about something he has done, her

audience applies what they know of him and of the situation to make a more specific attribution and then moves to predictions or explanations on the basis of that more specific thought. This too is a very general phenomenon: *When we attribute states to people using the folk psychological vocabulary, the content of the attribution is filled in by the audience in accordance with factors from outside that vocabulary*.

These three observations pull in the same direction. The vocabulary of folk psychology is a unifying point, a drop of glue, at which many disparate not essentially psychological capacities are brought together to give us a grasp of motive and action. Sometimes these other capacities do most of the work of predicting or explaining. (Usually they do, I think, but the point is obviously open to controversy.) Very rarely can the vocabulary be deployed in a self-sufficient way.

That unifying function is a deep and important one, though. It allows the disparate capacities to be linked and deployed together. And, also, its presence allows them to be rehearsed in application to other human beings, which is not at the centre of their natural range of functions. A good example of this, to be discussed in the next section, is the use of conditional thinking, primarily an action-planning capacity, to imagine the choices other people are likely to make in possible situations. These applications of capacities beyond their instinctive domains have to be learned, and we would not learn them, at any rate would not learn them in anything like the way we do, without the demand for attributions of states of mind and thought out rationalizations of motive and action in a specific and limited vocabulary. So, summarizing in a way that brings out both the centrality and the limits of the role that the vocabulary plays: Learning how to use the folk psychological vocabulary scaffolds the development and application of a host of other skills, all of which can be essential when we predict, explain, anticipate, or interact with others. (See Chapter 11 of Sterelny 2003, and Dan Hutto's contribution to this volume.)

12.4. LEARNING TO IMAGINE

For many philosophers and psychologists the use of the folk psychological vocabulary is not at the heart of folk psychology. Many writers have postulated a level of thinking about states of mind that cannot be easily expressed in ordinary language. So have I, in this paper, but while the thinking I have postulated is varied and in its non-verbal form not specific to the understanding of other people, for many writers there is a mind-specific domain of conceptual thinking, distinct from the manipulation of the surface vocabulary of folk psychology. (See Gopnick and Meltzoff 1997.) One form this can take is that of a theory of motive and action, some elements of which are innate and others of which develop in childhood, in terms of which we understand one another.

This account has never been refuted, but after its initial promise no one has produced a detailed version which explains any hitherto unexplained data. No one has said what the theory is, what its assumptions and rules are. And when philosophers try to describe the theory it collapses into the theory of rational action, which is indeed a coherent theory, but not implicit and not a theory of human

psychology. It is a theory that can be applied to human psychology, if many adjustments and provisos are added, but the adjustments and provisos have never been stated systematically, let alone presented in a way that it is plausible that they are part of something specific to the human understanding of other humans that slots into place in the first few years of life.

Other accounts concern "simulation": understanding another's mental processes by undergoing similar processes oneself. There is a large variety of such accounts, and some of the best known of them describe capacities that are clearly adaptations for psychological purposes of basically non-psychological skills. I shall briefly discuss two relevant skills. (The distinction between theory theory accounts and simulationist accounts is too crude: see the introduction to Davies and Stone 1995. Nichols and Stich (2004) speak helpfully of information-rich and information-poor accounts.)

In the cases to which off-line simulation accounts, pioneered by Robert Gordon, apply best, one person anticipates the action of another by activating her own decision-making processes as applied not to her actual situation but that of the other person, taking the output of this process not as a decision for herself but as a prediction about the other. (See Gordon 1988. Gordon has moved on from very simple simulationism now, see Gordon 1995.) Suppose you are trying to catch someone descending some stairs to a hotel lobby. At the bottom of the stairs she can go to the street through the door to the right or the left. You see obstacles between the stairs and the left door and so, without consciously thinking about it, you go towards the right door expecting that that is the way she will go.

How could someone learn to apply their own decision-making thinking to the actions of another? A natural suggestion, first made by Gregory Currie (1995), is that the connecting link is conditional thinking. That is, thinking in which one decides what one would do if various situations were to occur. This has an obvious utility: If the best response in the unlikely event of a kitchen fire is to first try to douse it then it makes sense to buy an extinguisher for oil-based fires now. Conditional thinking requires that one feed hypothetical inputs into one's decision making processes, and that one disconnect the output of the process from actual action. (The end process of thinking out what you would do if the chip pan were to burst into flames is not to spray with the extinguisher now, at the empty stove.) So the only further adaptation to get off-line simulation is the replacement of the input of one's own hypothetical situation with that of another (hypothetical or actual).

If one wanted to make off-line simulation into an all-purpose mind-understanding tool one would have to consider many more refinements. Some of one's own idiosyncrasies and many of one's own desires will often have to be subtracted, for many predictive and explanatory tasks, and characteristics of the other's thinking may have to be worked into the simulated decision-making. But it is not my aim to make it into an all-purpose tool. My aim is to show how a process that is useful in a limited range of situations, and very plausible as a means that we do employ in that limited range, can be understood as an adaptation for psychological ends of a capacity whose primary employment is quite different.

A rather different capacity that complements off-line simulation is that of perspective-taking. This is a general term for a capacity that begins with an infant's tracking the direction of gaze of a care-giver. This seems to be an innate human trait, not shared by our close relatives. (Though dogs are said to have something analogous, while wolves do not.) It has obvious application in pre-verbal social life and is thought by some developmental psychologists to play an important role in the development of social skills and of language. (To understand what an adult is referring to when using words a child has not heard before, it helps immensely to know what the adult is looking at.) As this capacity develops and becomes more sophisticated it grows, no doubt with help from other capacities, into the ability to know what things look like from another person's perspective, and then to less literally visual applications in appreciating another person's point of view on a situation.

Basic perspective-taking can support simple attributions of visual information to others. It allows one to know what another person can see. (Consider small children playing hide and seek: One thing they learn from this is to imagine what someone walking along a particular route will be able to see.) And these simple attributions can combine with simple social routines in the absence of any explicit folk psychological thinking. (Consider small children playing tag: You don't want the other person to touch you and so you want to know which way they will lunge and to do this you have to be aware of which way they are looking.) Moreover, basic perspective-taking is required for all but the most trivial applications of off-line simulation. As I described the example of the person waiting to intercept the other person coming down the stairs, the interceptor assumed that the descender could see the obstacles on the way to the left door. But it could be that the descender could not see the obstacles, and the interceptor was in a position to see that she could not. And then he should not anticipate her going to the right door. In modelling another person's decision-making with your own you have to feed into it the information that is actually available to the other person. Knowing what information this is, is sometimes so simple a business, for us humans to whom it comes naturally, that we can fail to see that it is a skill that has to be present and is sometimes so challenging that without a very advanced form of perspective-taking any attempt at simulation will be sure to fail. (See Morton (2006), and Chapter 5 of Morton 2003.)

Perspective-taking and off-line simulation illustrate my central point. We start with two non-psychological skills, gaze-following and conditional thinking, each of which can be adapted to serve a basic psychological function. Together, though, they form a more powerful predictive and explanatory device and fit together so easily that we might at first think we were dealing with part of a specific mind-appreciating module.

12.5. CONCLUSION: AS REAL AS BEARS

Imagine a set of circles on a plane. There are many overlaps between the circles, and if we attend to one such overlap we can see the circles that it is part of as peripheral extensions of it. But of course any of these circles could also be seen

as the peripheral extension of a quite different overlap, just as the overlap we first focused on could be seen as just one of many regions defined by the overlapping circles.

The circles are the fundamental human capacities. The overlaps are the many varied abilities we acquire by combining and adapting them: argument, public speaking, bicycle riding, violin playing, football, folk psychology. Some are more important than others. The overlap that constitutes folk psychology is no doubt an important one. But many of its overlaps with fundamental capacities are also parts of other derived capacities, some equally important even though we have not found standard labels for them. One could imagine a philosophy on which the ability to solve strategic problems (folk game theory) and the ability to feel sympathy for others were seen as important abilities underlying basic features of human life. Then the overlap that covers part of each of those, and others, that we call folk psychology might seem like an arbitrary and unnatural conglomeration.

This way of putting it obscures the difference between a weaker and a stronger possibility, though. The weaker possibility is that there is a fixed list of human capacities and that in all people there is a combination of these capacities, adapted in roughly the same ways and used in roughly the same proportions, that is used for everyday psychological explanation. The stronger possibility is that because individual variations in capacity and development the combination that one person uses to understand and predict others is significantly different from that used by another. We don't know which of these is nearer to the truth.

Suppose for the sake of argument that the weaker possibility is right, since it gives the greater solidity to folk psychology. Does it follow that folk psychology is real, that we can say "human beings have a range of abilities, and among them is the ability to anticipate actions in strategic situations, sense what emotions another is feeling, make decisions on other people's behalf, and the like"? Well, if the line explored in this chapter is right, it might be rather like saying "human beings have a range of abilities, and among them is the ability to ride unicycles and to play the oboe". Yes, each of these components exists, so in a way their combination exists, as real as the object consisting of the top 5 m of the Eiffel tower and Julius Caesar's left femur. Or, to use an analogy more like the one that began this section, folk psychology is as real as a constellation. The Great and Little bears, for example, are patterns in the sky that strike us as salient. Each of them, though is composed of stars many light years apart from one another, not related by origin or gravitation, and not forming any similar pattern seen from elsewhere in the universe. If the bears are astronomically real then folk psychology is an equally real part of human cognitive equipment.

REFERENCES

Bermúdez J (2004) The domain of folk psychology. In: O'Hear A (ed) Mind and persons. Cambridge University Press, Cambridge.

Camerer C (2003) Behavioural game theory. Princeton University Press, Princeton, NJ.

Currie G (1995) Imagination and simulation: aesthetics meets cognitive science. In: Davies M, Stone T (eds) Mental simulation: evaluations and applications. Blackwell, Oxford.
Fischer T (1993) Listed for trial. Granta 43:48–65.
Goldie P (1999) How we think of others' emotions. Mind Lang 14:394–423.
Gopnik A, Meltzoff AN (1997) Words, thoughts, and Theories, MIT Press.
Gordon R (1988) The structure of the emotions. Cambridge University Press, Cambridge.
Gordon R (1995) Simulation without introspection, or inference from me to you. In: Davies M, Stone T (eds) Mental simulation: applications and evaluations. Blackwell, Oxford.
Kusch M (1999) Psychological knowledge: a social history and philosophy. Routledge, London.
Morton A (2003) The importance of being understood: folk psychology as ethics. Routledge, London.
Morton A (forthcoming-a) Folk psychology. In: McGlaughlin B (ed) The Oxford companion to the philosophy of mind. Oxford University Press, Oxford.
Morton (1996) Folk psychology is not a predictive device. Mind 105:1–19.
Morton A (2006) Imagination and misimagination. In: Nichols S (ed.) The architecture of the imagination: new essays on pretense, possibility, and fiction. Oxford University Press, Oxford.
Nichols S, Stich SP (2004). Mindreading. Oxford University Press, Oxford.
Sterelny K (2003) Thought in a hostile world. Blackwell, Oxford.

MATTHEW RATCLIFFE

13. FROM FOLK PSYCHOLOGY TO COMMONSENSE

13.1. INTRODUCTION

The term 'folk psychology' is sometimes employed to mean 'our everyday, commonsense understanding of others' but can also be used to refer to the more specific view that everyday interpersonal understanding is enabled by a theory, rather than by simulation. The 'theory' theory is controversial. However, an account of 'commonsense' or 'folk' psychology, in the former sense of the term, is routinely accepted, according to which its central element is the attribution of intentional states, principally beliefs and desires, in order to predict and explain behaviour. For example, Stich and Ravenscroft (1996) observe that "ordinary folk certainly don't take themselves to be invoking a theory when they use intentional terms to explain other people's behaviour" (p. 117) and go on to claim that there are "various possible interpretations of the assumption that beliefs, desires, and other commonsense mental states are posits of a folk theory of the mind" (p. 124). Although such remarks indicate that the 'theory' theory is not part of commonsense, they also imply that the attribution of intentional states to explain behaviour is.

Stich and Ravenscroft distinguish two different interpretations of the claim that folk psychology is a 'theory'. It might be an internal structure that facilitates commonsense psychology or, alternatively, an external systematisation of folk psychological platitudes imposed by philosophers and others. So there are at least three different senses of 'folk psychology': (a) everyday talk and thought about beliefs, desires and other mental states, (b) the internal cognitive structure that facilitates it, and (c) a structure imposed upon it from the outside. My focus here will be on (a), which is presupposed by both (b) and (c). It is generally agreed by participants in the theory-simulation debate that 'folk psychology', understood in this broad sense, is not just a constituent of social ability but a central enabling condition for all social life. For example, Wellman (1990, p. 1) asserts that "an understanding of the mind is also fundamental to an understanding of the social world" and that this understanding consists of an ability to "explain our own and others' actions mentalistically, that is, in terms of the wishes, hopes, beliefs, plans, and intentions of the actor" (p. 8). And Churchland (1998, p. 3) goes so far as to state that folk psychology "embodies our baseline understanding of the cognitive, affective, and purposive nature of other people".[1]

How are such claims supported? Assertions and assumptions about the nature and centrality of belief-desire psychology appear frequently, without explanation or justification. So perhaps the answer is that little effort is required to illuminate the structure of 'commonsense psychology'. Most of the recent literature takes an appreciation of what we *do* for granted and focuses instead on the philosophical

D. D. Hutto and M. Ratcliffe (eds.), Folk Psychology Re-Assessed, 223–243.

problems that arise when it comes to explaining *how we do it*. Some claim that we employ a largely tacit theory; a systematically organised, domain-specific body of conceptual knowledge, embodied in the brain. Others claim that a large part of the burden is taken up by simulation; one uses one's own psychological mechanisms to model those of others and predicts their behaviour by working out what one would do in a similar environment or psychological state. There are also a number of hybrid theories, incorporating elements from both. But, despite numerous points of disagreement, most accounts start off with the same explanandum: *everyday interpersonal understanding, construed as the attribution of intentional states in order to predict and explain behaviour*. (Hereafter, I will refer to the conjunction of the claims that (a) we have a 'commonsense' or 'folk' psychology and (b) its main ingredient is an ability to attribute intentional states in order to predict and explain behaviour, as FP.)[2]

In what follows, I will challenge the orthodox account of FP in three ways. First of all, I will focus on the claim that certain abilities or concepts are part of our social 'commonsense', of a 'folk' view, and will show that proponents of FP have neglected to clarify what they mean by 'commonsense'. As a result, they fail to demarcate commonsense views from debatable philosophical positions. Then I will investigate the alleged scope of FP. Like other contributors to this volume, including Morton, I will argue that there are different kinds of social situation, which demand a number of different social skills. I will suggest that FP is not needed at all in some social situations and that its relevance to others is debatable. Despite this, participants in the theory-simulation debate have neglected to offer a clear account of its scope, the exception being those who have asserted, without argument, that it is the source of all social life. Finally, I will turn to the main ingredients of FP, focusing on belief. I will show that the term 'belief', as employed by proponents of FP, encompasses many different psychological states, in addition to features of situations. Although everyday talk allows us to intuit the differences between these states, examples of belief-desire explanation, of the kind routinely offered by proponents of FP, do not. Consequently, 'belief', as the term is employed in much of the folk psychology literature, is an abstract placeholder for a variety of psychological states that feature in everyday explanations of action. I conclude by suggesting that FP has no psychological reality and is instead an abstract philosophical systematisation of social life, the utility of which is unclear.

13.2. FOLK DUALISM

Some of the problems involved in stating a 'folk' or 'commonsense' view can be clearly seen in Paul Bloom's recent claim that we are commonsense Cartesian dualists. According to Bloom, "dualistic thinking comes naturally to us. We have two distinct ways of seeing the world: as containing bodies and as containing souls" (Bloom 2004, p. xii). He suggests that humans have innate tendencies towards dualism, which are evident in babies before they can even speak or walk. For example, when an object they are watching stops moving, babies will get bored

and stop attending to it. But when a person stops moving, they will more often become distressed, demonstrating a difference between their responsiveness to people and to inanimate things (Bloom 2004, Chap. 1). These tendencies develop into the commonsense view of adults, which can be inferred from what they say. Explicitly stated belief in the afterlife, for instance, is commonplace (Bloom 2004, pp. 205–208).

However, neither source of evidence provides adequate support for the claim that substance dualism comes naturally to us. Although someone might claim to believe in disembodied minds that survive death, her view will most likely become considerably murkier to both herself and her interrogator upon more detailed questioning. Consider the following questions:

1. Will you have a spatial and temporal location after death?
2. Will you be able to see and hear?
3. Will you be able to act and, if so, will there be any constraints to your ability to act?
4. Will you be able to communicate individually with other disembodied spirits?
5. Will you have a specific point of view, from which you experience things?

If someone answers 'yes' to all or most of these, then his 'disembodied' spirit looks to be tacitly embodied and if he answers 'I don't know', then it is unclear what he really thinks about disembodied existence. I am not sure exactly how most people would respond to such questions but answers will no doubt be more varied and hesitant than they are to the question 'could you exist without your body?'[3]

The claim that young children are dualists appears to be better supported. Wellman (1990, p. 50), like Bloom, argues that humans are commonsense dualists and, in support of the view, offers a series of carefully considered experiments designed to tease out folk ontological distinctions from the verbal reports of young children. According to Wellman, "natural language research suggests that children understand some appropriate distinctions between mental and physical entities and events towards the end of the third year of life" (Wellman 1990, p. 23). In making such claims, he assumes an understanding of what adults mean when they employ mentalistic terms and addresses only the question of whether the same concepts are grasped by children. Hence both Bloom and Wellman take for granted that ontological commitments can be read off adult everyday language with little effort. They can't. Consider what Strawson (1959, p. 9) calls 'descriptive metaphysics', a form of philosophical enquiry that seeks to make explicit the structure of our actual conceptual scheme. Descriptive metaphysics is not simply a matter of reading metaphysics straight off everyday utterances. People are seldom clear or explicit about what they are committed to and the task of mapping our conceptual scheme is thus a difficult one, requiring complicated philosophical arguments and exercises of the imagination. Although Bloom (2004, p. 195) observes that "our bodies are described as our possessions" and takes this to be evidence for dualism, he neglects to note that we also refer to our minds and mental states as our possessions. The 'I', 'you', 'he' or 'she' of everyday language is ascribed both mental and physical predicates. For example, 'I' am six feet tall and 'I' am thinking about philosophy.

This is one reason why Strawson (1959, Chap. 3) claims that Cartesian dualism is not an ingredient of our actual conceptual scheme. The referent here, according to Strawson, is neither a mind nor a body but a 'person' and the concept of a 'person' is a primitive, unanalysable constituent of our conceptual scheme. Cartesian dualism incorporates a revision of that scheme, whereby 'I' either refers to both of two distinct substances or to just one of them. Everyday Anglophone discourse neither distinguishes a corporeal 'I' from a non-corporeal 'I' nor incorporates a distinction between 'I am thinking' and 'I am six feet tall' to indicate that only the former really involves attribution of a characteristic to an 'I'.

The evidence provided by Bloom and Wellman is compatible with Cartesianism *and* Strawsonianism; that we distinguish in various ways between objects and people is acknowledged by both views.[4] But the problem is not just that 'folk dualism' is underdetermined by the evidence and that other views, such as that of Strawson, are also candidates. Of far greater concern is the lack of any distinction between the project of descriptive metaphysics and the task of describing a 'folk' or 'commonsense' view. Both relate to our actual conceptual scheme. The difference is that Strawson recognises the difficulty of articulating it and treats the result as a philosophical position. So to simply assert 'the commonsense view' is to make a philosophical claim without doing the philosophical work. It is not clear that there is such a thing as a 'folk metaphysics', to be set apart from various conflicting philosophical accounts and arguments.

Bloom's claim that we are natural born dualists also fails to distinguish conceptual from non-conceptual abilities. Even if babies do behave in a way that is consistent with their being dualists, they do not have a complicated conceptual understanding of minds and bodies. Adult dualism, in contrast, is construed as a conceptual structure that is evident in our various utterances. However, words and behaviour often part company and 'practical commonsense' need not cohere with 'verbal commonsense' (Greenwood 1999). It is conceivable that someone might act in a way that accords with dualism, whilst verbally rejecting it, and vice versa.[5] Hence 'commonsense' is ambiguous; infant and adult dualism could be quite different things. In addition, people might talk or act in a way that is consistent with dualism only in some social situations. Further ambiguities arise with respect to verbal commonsense, as it is not clear which instances of verbal behaviour should be taken into account or by whom. Casual, uninterested responses to questions about minds and bodies, such as 'yeah', 'maybe', 'dunno', 'guess so' and 'don't care' are presumably to be discarded. And Cartesian dualism cannot simply be read off richer conversational narratives. 'Hey, you're right – I got soul man!' is not an explicit assertion of dualism. Careful reflection is surely required in order to decide on the kinds of discourse to be scrutinised and then to make explicit the concepts underlying that discourse. But this is a philosophical exercise and the result will be a philosophical position as open to debate as any other. So we return to the earlier problem. The claim that we are innate, commonsense substance dualists is vague and unsubstantiated; a philosophical claim made without philosophical thought.

One might reply that, although commonsense Cartesianism is dubious, the central aim of both Wellman (1990) and Bloom (2004) is to clarify an understanding of the difference between mental and physical states that all normal humans above a certain age possess. Both authors ultimately place much more emphasis on 'mindreading' than 'dualism'. In other words, their primary emphasis is on our ability to attribute intentional states in order to predict and explain behaviour (FP). This is dissociable from the case for dualism and so not susceptible to the above criticisms. However, I suspect that the same problems arise for FP as for folk dualism, even though the latter is an easy position to criticise, whilst the former is accepted without question in many areas of philosophical and scientific enquiry.

13.3. ASKING THE FOLK

What does it mean to claim that 'attribution of intentional states in order to predict and explain behaviour' constitutes a 'commonsense' or 'folk' psychology? One answer is that we're all well aware that this is what we do; unlike folk dualism, it really just is commonsense. I asked a tutorial group of second-year philosophy undergraduates what they thought the central ingredients of interpersonal understanding were. They were halfway through a philosophy of mind course at the time and had already been given three lectures on intentionality but were not yet familiar with the term 'folk psychology'. The exercise was not very successful, given that most of them had trouble understanding the question. In attempting to clarify it, I found myself increasingly lapsing into debatable philosophical positions, which my students would then presuppose in their answers. Now if the question were worded as follows, most answers would doubtless conform to FP:

> Actions are caused by entities inside people's heads called beliefs and desires. Different combinations of beliefs and desires cause different actions. Given this, what do you have to do in order to predict which action another person will perform?

But this would be to smuggle the answer into the question. Of course, there is no uniquely appropriate or wholly neutral way of phrasing the question. Nevertheless, I think it can be formulated in a way that is not too vague or general to facilitate a clear response but also not so specific that the presuppositions of FP or some other equally specific view are incorporated into it. With the following year's groups, I settled on:

> What is central to your understanding of others? To put it another way, understanding or interacting with another person is very different from understanding or interacting with a rock. What does that difference consist in? Please state your intuitive or commonsense view rather than stating philosophical positions or engaging in philosophical argument. Write up to half a side of A4 and return it to me at next week's tutorial.

Twenty five responses were handed in, which listed a diverse range of factors. Here is a selection:

"Can understand me in conversation."
"Can detect their emotions through facial expression and body language."

"Can relate to my plight."
"An understanding of why we do certain things in certain situations."
"Difficult to say exactly […] lots of things."
"Empathy."
"Knowledge of the individual."
"Their relationship to ourselves."
"The soul knows itself, and it knows others, for all souls come from the one."
"You can interact with others and see the way people react to things."
"They act similarly to us."
"How they respond to me."
"Being able to care for their well-being."
"It's when you're in Kingsgate bar with some friends, having a drink, when Franz Ferdinand comes on the radio and you notice that everyone in the bar has started nodding along to it…and so have you."
"Reactions and engagement – able to interact in the world."
"The same type of thing – the same species of entity – as myself."
"Others have the same facial movements as me and show emotion via the same route."
"We don't understand everyone! e.g. psychos, scientists."[6]

The term 'belief' appeared twice in total and 'desire' appeared once, as did 'prediction'. 'Explanation' was not mentioned at all.[7] Given this, it would appear that FP is not a 'commonsensical' view of what social life most centrally involves. Indeed, the diversity of rather vague responses suggests that there is no explicit, shared, commonsense conception of everyday social understanding. It might be objected that my students simply struggled to articulate their grasp of FP or deemed it too obvious to mention. However, when I offered it as one possible answer, I was not met with any sign of recognition or eager assent. Many of the students later came to accept FP, after five lectures on folk psychology, theory of mind and simulation. But adoption of a philosophical viewpoint, whether due to conformity, indoctrination or rational deliberation, does not amount to clarification of a commonsense view that was held prior to one's engagement with academic philosophy.[8]

Although these observations certainly do not constitute a rigorous scientific experiment, they are, I think, sufficient to illustrate that FP is not the outcome of commonsense reflection on social life. The term 'commonsense' could be employed to refer to conceptual or non-conceptual abilities. One possibility is that social commonsense is largely a matter of the latter and is thus not something that we should expect people to articulate easily. However, it is quite clear that proponents of FP take it to centrally involve the possession and application of concepts. As Churchland puts it, FP is "the prescientific, commonsense conceptual framework that all normally socialized human beings deploy" (Churchland 1998, p. 3). This view is not restricted to those who take folk psychology to depend upon a theory, a systematically ordered, largely tacit body of conceptual knowledge. Theory theories postulate a lot more conceptual knowledge than simulation theories. However,

concepts such as 'belief' and 'desire' are claimed to be evident in everyday platitudes about mental states and behaviour, which are presupposed by both kinds of theory. They are not the tacit posits of a theory. Indeed, folk psychological concepts can't be wholly tacit if they are routinely incorporated into *explanations* of behaviour. But the comments of my students also indicate that FP is not a conceptual understanding whose basic structure is routinely articulated. Thus FP must be a conceptual ability that we routinely employ but which does not easily spring to mind when we reflect upon what we do. Hence social ability must be divided into *at least* three distinct categories, with FP falling under 2:

1. Non-conceptual, practical social abilities.
2. Conceptual social abilities.
3. Ability to reflect verbally on 1 and/or 2.

How we actually think and behave in social situations need not always correspond to what we say we think and do in such situations. Perhaps, therefore, FP is social commonsense interpreted from an appropriate standpoint of expertise. One possibility is that it is a description of 'everyday psychology' obtained through *scientific* studies of our *conceptual* social abilities. However, many discussions suggest that this is not the case and that commonsense concepts are not generally revealed through scientific investigation. For example, the original false belief task (Wimmer and Perner 1983) and its many subsequent variants *presuppose* that we employ FP and instead enquire as to when aspects of FP ability, principally the ability to appreciate that another person has a belief that differs from one's own, develop (Ratcliffe 2007, Chaps 2 and 4). Other discussions also indicate that commonsense concepts are not discovered through scientific study. For example, MacDonald (2002, p. 467) discusses intuitions concerning first-person knowledge of mental states and refers to what "commonsense tells us" about knowledge of our own mental states, thus suggesting that commonsense intuition, rather than scientific study, is the means of access to at least some aspects of our folk view.

Hence the rationale for describing the core of social understanding as 'the attribution of intentional states to predict and explain behaviour' remains unclear. Both theory and simulation theories are able to explain why we are not fully aware of *how* we manage to ascribe intentional states to others. For the theory theorist, there is a complex and largely *tacit* conceptual structure underlying the ability to attribute intentional states. For the simulation theorist, attributing the outputs of one's own cognitive processes to others need not involve much or even any knowledge of *how* the processes that produce those outputs actually work. However, that we *do* attribute beliefs and desires in a way that is central to social life is something that most versions of both theories assume from the outset as an explanandum. Scientific work is generally concerned with differentiating the two accounts and investigating the developmental trajectory of FP, rather than with the question of whether FP itself is an adequate description of the achievements central to human social life. Even if scientific evidence were invoked to defend FP, one would be faced with the same problem as 'commonsense dualism'. The explication of conceptual structures that we take for granted in everyday life is a philosophical task that cannot be

avoided by referring to commonsense or what the folk think. And a conceptual scheme cannot simply be read off experimental results concerning verbal and non-verbal behaviour. Indeed, our most deeply entrenched concepts will most likely be presupposed as a lens through which scientific results are interpreted, rather than discovered through them.

All this suggests that FP is in fact something that has been made explicit through philosophical reflection. It is a philosophical account of the structure of everyday interpersonal understanding, rather than a commonsense description of something that we do. The labels 'folk' and 'commonsense' are therefore misleading. However, perhaps a very simple philosophical argument, acceptable to all, would be enough to illustrate the nature and centrality of FP.[9] It might then be 'close enough' to commonsense to merit the label. In the remainder of the discussion, I will suggest that this is not the case. In Section 13.4, I will ask *where* FP (construed as the attribution of intentional states in order to predict and explain behaviour) is supposed to apply and will suggest that its *scope* is far from clear. In Section 13.5, I will take the argument a step further. Even if we do assume that FP is applied in some contexts, we find that various illustrations of FP at work, which have been offered by its proponents, are quite abstract and fail to distinguish a range of different psychological states and other factors that we readily distinguish in everyday life. Hence they are far removed from how interpersonal understanding actually seems to work.

13.4. THE SCOPE OF FOLK PSYCHOLOGY?

The description 'X attributes beliefs and desires to Y in order to predict and explain her behaviour' is rather abstract and removed from the complexities of actual social situations. It goes without saying that we do not observe others in the total absence of social context and then attribute various internal states to them from a neutral, wholly detached perspective. Although it is often claimed that FP is a "cornerstone" of all social life (Langdon et al. 2002, p. 74), social life is not an amorphous whole. It is differentiated into various types of situation and these appear to involve various different kinds of understanding. The question therefore arises as to where and how FP is supposed to apply.

Schutz (1967) makes a number of helpful distinctions between different kinds of social understanding. For example, the conceptualised, detached understanding of social behaviour that a social scientist seeks is, he claims, quite different from the social experience of two engaged conversational participants:

> My [conversational] partner and I, for instance, have intimate and rich experience of each other as we talk together, whereas we both appear to a detached observer in an aura of 'flatness' and 'anonymity' (p. 8)

In conversation, Schutz claims that the other person is encountered through a stance that he calls a "Thou-orientation" (p. 153), through which she appears as a 'you', as opposed to a 'he', 'she' or 'it' to be scrutinised and conceptualised from a theoretical distance. A 'Thou-orientation' is a way of experiencing, rather

than a way of reflecting upon experience. Through it, 'you' are encountered "as a field of expression on which I can 'watch' the flow of your lived experiences" (p. 117). Hence there is a distinction between third-person scientific scrutiny and second-person experience.[10] Schutz adds that social encounters with a 'he', 'she' or 'it', though they do not incorporate interaction to the same extent as an 'I-you' exchange, can also be a form of social *experience*, distinct from third-person *theorising* about others. As such, they still incorporate a more general kind of personal standpoint, which he calls an "Other-orientation" (p. 153). Schutz further divides everyday social life into three different categories. There are my consociates, those I experience as 'Thou' and with whom I have social relationships. More distant are my 'contemporaries', who are often not experienced as persons but understood as impersonal social types or "ideal types" (p. 185), such as 'banker', 'builder', 'butler' and 'bar person'. More distant still are my predecessors, with whom I can have no social interaction.

Although the specific distinctions made by Schutz are debatable, they do serve to illustrate that social understanding and interaction come in several different forms. With this in mind, to which cases is FP applicable? Consider an example of interaction with contemporaries that is tightly bound up with an understanding of equipment. One descends an escalator towards the Northern Line Northbound platform on the London Underground, as one does every day at more or less the same time. In order to get down the escalator, buy a ticket, walk to the platform and get on the right train, one has to negotiate one's way around hundreds of people and coordinate one's behaviour with theirs. In routine cases like this, FP does not seem to have a role. The claim that a person navigating the London Underground with thousands of others is simultaneously attributing beliefs and desires to everyone she coordinates with or might have to coordinate with implies an enormous burden on her cognitive resources. Furthermore, it is an unnecessary burden. A shared understanding of how standardised pieces of equipment, such as escalators, ticket machines, platforms, trains and signs, function is at the same time an understanding of what people do in this kind of situation, regardless of who they are. Mental states do not need to be assigned, as the assumption that others will do 'what one does' in this kind of equipmentally configured environment is usually sufficient.[11] So it would seem that FP does not play a central part in certain routine social interactions with contemporaries. One might object that these are not 'genuine' social interactions. However, this would be question begging if genuine social situations are simply defined as 'those in which FP is deployed'. Although I may not attribute intentional states to the thousands of people I coordinate with on the London underground system, I am still aware of them in some sense *as people*, as quite different from inanimate objects and from the rodents on the tracks.

What about more personal encounters between consociates? Take the example of an everyday conversation. Does FP play a central role in facilitating mutual understanding? It is not clear that it does. As Goffman (1982) observes, conversation incorporates a complex web of conventions, which serves to regulate interaction and constrain mutual interpretation. There are standards of etiquette, ceremonial rules,

various kinds of perceived transgression and established techniques for restoring order. There are general expectations with regard to demeanour and there are deference structures appropriate to interactions with people of certain social statuses in certain contexts. It can be added that the evolving conversational narrative itself plays a considerable role in constraining interpretation (Ratcliffe 2005a). When a topic is mentioned and met with interest, it is built upon by both parties, through an interplay of word, gesture, tone, expression and feeling. A dynamic context of shared understanding and agreement forms, which structures ongoing interaction. Attribution of mental states from a detached, observational standpoint is hardly typical. One acts, reacts and comes to understand the other person through one's own actions and her responses, which are interpreted within the context of the shared conversational narrative. In addition, past conversations with specific individuals constrain future encounters, as person-specific knowledge, mutual expectations and shared routines all develop. So there is a lot more than FP going on in conversations. Furthermore, interactive construction of a context for mutual interpretation is very different from the attribution of beliefs and desires in the absence of shared context. Hence FP is, at most, only part of the story and, even if this much were conceded, it is not at all clear how FP ability is supposed to be integrated into the structure of interaction.[12]

It might be that everyday conversations where little is at stake for either party are not the kinds of environment in which the role of FP is most evident. The need to work out the underlying mental states of others is much more pressing in strategic interactions and one-off rather than routine situations, where obtaining a desirable outcome for oneself depends upon one's ability to work out what others will do. Those others are unlikely to simply tell you what they will do, given that a desirable outcome for them will often depend on you not getting what you want.[13] However, the role claimed for FP is unclear even in these cases. Consider a well-known example from football (soccer). In the 1996 European Championships, Paul Gascoigne scored the second goal for England against Scotland by sprinting forward without support, lobbing the ball over the head of Scottish defender Colin Hendry, running round him and then kicking it into the corner of the net past goalkeeper Andy Goran. This was a strategic situation in which the preferred outcome for Gascoigne depended upon others not getting what they wanted. And his ability to anticipate complicated behaviours on the part of others and coordinate with them in a structured, ongoing, open-ended fashion was clearly in evidence. But was this display of genius an instance of particularly good FP thinking?

I mean no disrespect in suggesting that Gascoigne is not an all-round Machiavellian social genius. His remarkable feat was the product of skills specific to the pitch. It is possible that his exceptional skill at football enabled him to employ his FP capacities more effectively on the pitch than in other situations but this still entails that domain-specific skills comprise a substantial component of social ability. In addition, the speed of Gascoigne's responses suggests that he had no time for deliberation. If neurobiological findings reported by Libet (2004) are accepted, there would not have been sufficient time between stimulus and response for him

to even be aware of some of his responses before they happened. This should come as no surprise. As Libet observes, "great athletes, in general, are those who can let their unconscious mind take over without interference from the conscious mind. Athletes tell us that if they try 'to think' (become aware) of immediate responses, they become less successful" (Libet 2004, p. 111). So either there is an unconscious belief-desire psychology unfolding at an amazing speed or something else is going on. Expert sportspeople and game players, when asked why they responded in a particular way, will often report that they 'just saw it' or 'did it', rather than offering a complicated narrative concerning their ability to attribute internal mental states. Gascoigne comments, with respect to Hendry, "I just knew where he'd be, when he'd commit himself, so I knew what to do" (Gascoigne 2004, p. 214). As Dreyfus and Dreyfus argue, expertise seems to involve attuning oneself to a situation in such a way that the situation itself drives and structures one's action; "once one has a skill one is solicited to act without needing to have in mind any goal at all" (Dreyfus and Dreyfus 1999, p. 111) and explicit, conceptual cognition drops out. So there is evidence to suggest that many strategic interactions in sport do not involve the deployment of FP. Now our FP advocate might simply state that FP was never intended to apply to this kind of fast-paced, skilled activity. However, this misses the point: The social abilities demonstrated by Gascoigne in the context of football are just as complicated and ingenious as those we might observe in other, quite different strategic social situations. So, if one can do without FP here, why not in those cases too? Even when one is not running around on a pitch or making split-second decisions, context-specific skills could well be driving the ability to interpret and interact with others.

One response on behalf of FP is that, although Gascoigne may not himself have been aware of employing FP at the time, FP is still quite clearly present in the way such achievements are talked about. However, I suggest that even this much should not be conceded. I have observed in this section that social life encompasses different kinds of situation and that it is not clear where FP is supposed to apply or how central its role is supposed to be. However, I will now suggest that, even if we restrict ourselves to the kinds of examples of FP in action that are offered by its proponents, some such examples turn out to be too abstract to facilitate an adequate understanding or explanation of action and others are irrelevant. I will focus on how the term 'belief' is employed in order to illustrate these shortcomings.

13.5. BELIEF

Most accounts of FP take the commonsense conception of 'belief' to be unitary and tend to offer something like the following:

> Believing is the mental attitude of conviction, the thought that something is true. It is the attitude that takes a description ("the car is green") as corresponding (to some degree) to a state of affairs in the world (Wellman 1990, p. 61).

However, I will argue here that a host of different states that we intuitively distinguish in everyday life are surreptitiously lumped together under the label 'belief' by

proponents of FP.[14] Of course, accounts of FP often explicitly restrict themselves to belief construed as a 'propositional attitude'. In other words, 'belief', in the FP sense, takes the form 'X believes that p', where p is a proposition, such as 'the Eiffel Tower is in Paris', 'it is raining', 'the supermarket is open' and so forth. However, I will suggest that, even if this is the case, the kinds of concrete examples of FP that are often appealed to are suggestive of a broader conception of belief. Furthermore, even if we do restrict ourselves to propositional attitudes, there are additional problems. Depending on how we understand 'propositional attitude', either we get a concept of belief that still accommodates a range of different states or one that is so narrow as to exclude almost all of the states that we ascribe in order to understand social action. In addition to all this, much of our everyday talk about action does not map onto the kind of examples offered on behalf of FP as 'platitudes'. Whereas FP is abstract and insensitive to broader social context, appreciation of social context plays a key role in facilitating most instances of interpersonal understanding. Social behaviour is not generally understood by attributing internal causes in the form of beliefs but by placing it in the context of familiar kinds of structured social situation.

Last year, my wife moved the kitchen bin from one side of the room to the other and informed me that she had done so. Shortly afterwards, I walked over to where the bin had previously been and dropped a banana skin on the floor. For several weeks, I continued to approach the bin's former location, rubbish in hand, before changing my trajectory at the last moment. Did I continue to believe that the bin was where it used to be? It would seem not; were someone to ask me where my kitchen bin was located, I would have answered without any hesitation that it was at the new location. However, surely my actions indicate that I believed otherwise? Now one might say that I believed the bin to be in location A but took my rubbish to B due to engrained habit, meaning a set of non-conceptual dispositions. However, it is difficult to draw a clear line between non-conceptual habits and conceptualised belief. Consider a second case. Two months ago, my bank sent me a new Visa card, accompanied by a new PIN number. I'd had the previous PIN number for ten years and, on the first few occasions when I used the new card, I typed in the old number 4256 rather than the new number. If asked by an appropriate authority, I would have reported that my PIN was **** and not 4256 but my actions were in conflict with this. Now I may have typed in the old number out of habit but, if so, habit starts to look far richer than a set of non-conceptual routines. In one instance where I pressed 4, 2, 5, 6, the cash machine was one I had never used before and I was unfamiliar with the way its buttons were configured. I surely had to think about what I was doing, recognise the numbers on the buttons as 4, 2, 5 and 6 and 'know' that I was pressing them. Yet I still typed in my old number. Did I therefore believe both 'p' and 'not p'? This doesn't seem to trouble everyday 'folk' and a simple telling of the story is enough to make the behaviour intelligible. Rather than reporting my beliefs, desires and habitual tendencies, it is enough that I describe the situation: "I received a new PIN after using the previous one for ten years and I still keep typing in the old one." I might say that I "keep thinking" I still have

the old one or that I'm "used to" the old one but it doesn't really matter precisely which psychological terms are used. I have described a familiar type of situation and that is usually enough.

An intuitive grasp of some kind of distinction between habitual actions and the tendency to make explicit utterances is clearly part of everyday life but it is not evident in many examples of 'folk psychological platitudes'. Consider the following:

> If a normal person is looking at a traffic light which changes from red to green, she usually comes to believe that it has changed from red to green (Nichols and Stich 1996, p. 126).
>
> …it is trivially easy to explain why John will carry his umbrella with him: it is because he *believes* it will rain and he *wants* to stay dry (Frith and Happé 1999, p. 2).

Are these cases more like my articulated 'belief' that my wife had moved the bin to B or my practical 'belief' that the bin was at A? The first could involve a habitual response, the unthinking movements of hand and foot that come into play whenever the lights change. The second is ambiguous. Consider the following:

1. As always, John switched off his alarm clock and got out of bed at 7:30. He dressed, ate breakfast, picked up his briefcase and umbrella and set off to work at the usual time of 8:30.
2. John opened the door and saw the unusually dark sky. He went back into the house and picked up his umbrella.

It is clear from these descriptions that quite different psychological states are at play in 1 and 2. 1 describes a routine. John picked up his umbrella because that's what he always does. In 2, John's picking up the umbrella is preceded by an explicit thought. Replacing them both with 'John's believing it will rain and not wanting to get wet' obscures the difference and detracts substantially from an understanding of his behaviour. There are further ambiguities evident in examples of so-called 'platitudes'. Consider the following:

> If you see a person running to catch up with a just-departing train, for example, you interpret the person as an intentional agent, who *believes* that there is a just-departing train, and who *wants* to get on it (Scholl and Leslie 1999, p. 131).

Although the terms 'believe' and 'want' are used in what looks like the same way as the umbrella case, very different situational and psychological predicaments could be involved, even if we take umbrella case 2, where 'it will rain' and 'I do not want to get wet' are explicitly entertained. To make this clearer, compare "Jane believes that the Eiffel Tower is in Paris" to "Jane believes that her train is about to leave". In the former case, Jane's 'belief' may well amount to no more than her disposition to affirm certain propositions relating to Paris and the Eiffel Tower, with no associated experiential qualities and no connection with any of her actions, save her explicit utterances. Thus it is a state best characterised in linguistic terms. However, Jane's believing that the train is about to leave is most likely an experiential state, in which experience and action are very closely tied together. The way that the train 'appears to Jane' is infused with emotion. Her experience of it is, at the same time, a reflection of her own predicament. Her running, the sight

of the train, the sound of the whistle, the shutting of the doors, a sense of urgency and a background of concerns and projects all blend together seamlessly. Just about any everyday description of the two cases would succeed in communicating the significant differences between them but stating that utterances and actions are caused by 'beliefs' and 'wants' or 'desires' does not. The two 'beliefs' seem to be quite different states, connected in very different ways to experience and action. Adding desires does not help either. Jane's believing that the train is about to leave and desiring to get on it may be quite different in nature to her believing the Eiffel Tower is in Paris and desiring to go there.

Everyday descriptions of people in situations generally succeed in distinguishing such psychological predicaments but do not, of course, amount to a rigorous analysis of what those differences consist of. Such analyses are a task for a philosophical psychology. But even in the absence of a comprehensive analysis of our various social concepts and abilities, it is apparent that everyday language tends to describe situational predicaments, rather than list 'beliefs', and that translating it into talk of 'belief' is inadequate because significant differences between kinds of psychological state are masked.

The FP conception of belief not only blurs the distinctions between various psychological states but also confuses psychological states with features of a person's situation. Consider a description of a commonplace situation: *There is a bus at the side of the road, by the bus stop, twenty yards in front of the zebra crossing. Across the road is a pub, a bank, a newsagent and a bookmaker.* Any such description will incorporate a host of references to the socially configured world, to tools, signs and artefacts more generally. Kinds of action and reasons for action are incorporated into an understanding of what such things are. For example, in perceiving a screwdriver, I do not conceptualise the object and *then* recognise it as something with which only *I* can drive screws. It is understood as something *for* a certain role and the concept 'screwdriver' thus serves to partially specify what *people* do with it, regardless of *who* those people are. In describing a situation as incorporating interrelated entities like buses, roads and crossings, I also incorporate an understanding of the kinds of action appropriate to that situation. This does much of the interpretive work required to understand what a particular person has done, is doing and will do. How? Do I simultaneously attribute billions of intentional states to each specific individual in a situation, such as "X believes that A is a bus; X knows what a hammer is; X thinks that pubs are for drinking beer in?" I do not. Taken to its extreme, this amounts to the suggestion that we concoct an entire experiential and cognitive world, different from our own, and attribute it to a person every time we try to interpret him. But all interpersonal understanding assumes that we are *in* some kind of common world with others and it is not just a world of space, time and matter. The world we generally presuppose as a shared backdrop for mutual interpretation is a world of buses, shops and pubs, in which certain kinds of activity are to be expected. I don't take *this world* to be something I experience and think about, and only then, having assigned several billion internal representations to you, something you experience and think about in a similar way. It is usually

assumed from the start that other people are competent participants *in* much the same social environment. It is enough to say that Jane moved the car because the light changed from red to green. Reasons for action are incorporated into the shared social environment. Stating that she 'believed it' does not add anything at all and it is not implied by everyday discourse.

Artefacts are not the only features of situations that specify kinds of action. The same point can be made with respect to what Schutz calls 'ideal types'; social roles or positions that are associated with certain kinds of behaviour. What FP takes to be an understanding of internal beliefs is often actually an external contextualisation, a description of a situation incorporating characteristic equipmental configurations and social roles. Consider the following:

> If I observe, or even hear about, a man tightening a nut, my first interpretive scheme will picture him as joining together two parts of an apparatus with a wrench. The further information that the event is taking place in an automobile factory permits me to place the operation within the total context of 'automobile' manufacturing. If I know in addition that the man is an auto worker, then I can assume a great deal about him, for instance, that he comes to work every morning and goes home every night, that he picks up his check every payday, and so on. I can then bring him into a wider context of meaning by applying to him the ideal type 'urban worker' or, more, specifically, 'Berlin worker of the year 1931'. And once I have established the fact that the man is a German and a Berliner, then all the corresponding interpretive schemes become applicable to him (Schutz 1967, pp. 192–193).

The activity of tightening the nut is understood by adding various layers of context, which involve increasingly general social roles and statuses, in addition to kinds of work and equipment.[15] But would advocates of FP really substitute something as uninformative as "he is tightening the nut because he believes he can tighten it and he desires it to be tighter" for the kind of rich contextualisation above? The answer is yes. Here's an example from Wellman (1990, p. 8):

> John's going to the store to buy groceries, for example, is explained by John's desire to eat and his belief that he can buy food at the grocery store. Similarly, I may decide to go to the grocery store rather than to the drugstore because I desire to get food, not pharmaceuticals, and I believe that food is found at the grocery store.

This pseudo-explanation has no relevance to the understanding of social action and encompasses all manner of situations and psychological predicaments. Consider the following candidates for why John went to the grocery store:

It was shopping day.
John was just passing by on his way back from work and thought he might as well pop in.
They had a special offer on tomatoes.
John had some spare change and he fancied an apple and a can of soda.
John is cooking for friends tomorrow and the shop won't be open then.
John is a health inspector and it's his job to check the quality of all food sold in the area and he's just bought a selection of foodstuffs to take back to the laboratory.
John was doing some shopping for the old lady down the road.

John wanted to keep a conversation going with Sandra. So he followed her into the grocery store and did some shopping himself so as not to make it too obvious.
John is a chef and he went in to buy some food for his restaurant.

Understanding John's action involves an appreciation of how it is incorporated into a situation, a project or the structure of a life. Reference to mental states will often have a role to play but these are knitted into a broader context and are understood through that context. Indiscriminate application of the term 'belief' is no substitute for a more refined contextualisation. So how is the term 'belief' actually used? Bruner (1990) suggests that it only enters into our descriptions of action when people deviate from what he calls the canonical narratives of a culture, an inventory of familiar situations, with familiar roles, norms of performance and patterns of activity attached to them. 'Belief', according to Bruner, is employed as part of a story whose function is to integrate unusual behaviour back into familiar social patterns and thus make it intelligible:

> The function of the story is to find an intentional state that mitigates or at least makes comprehensible a deviation from a canonical cultural pattern (pp. 49–50) …human beings, in interacting with one another, form a sense of the canonical and ordinary as a background against which to interpret and give narrative meaning to breaches in and deviations from 'normal' states of the human condition (p. 67).

However, various different states, such as sentential attitudes, habits and experiences, could all serve the same function in such stories. So the fact that 'belief' is sometimes used in such a way does not serve to single out something that is distinctive about 'belief'. Indeed it seems that the term is not used in a single, clear way and Bruner's account only encompasses one variant.[16] Here are a few others:

Astonishment: "I don't believe it!"
Disappointment: "I don't believe it!"
Incredulity: "I don't believe it!"
Anger: "I can't believe you did that"
Disgust: "I can't believe you did that!"
Forms of trust: "I believe you" or even "I believe in you."
Moral conviction: "I don't believe that's right."
Religious faith: "I believe."
Uncertainty: "I believe it is raining but I'm not sure."

Differences in what is meant by belief are obscured by so-called 'platitudes' about traffic lights, umbrellas and departing trains. What is meant by 'belief' on a particular occasion is more easily understood when an utterance is placed in context. For example, at the Labour Party Annual Conference on 29th September 2004, the British Prime Minister Tony Blair said the following about his decisions concerning the ongoing situation in Iraq:

> Do I know I'm right? Judgements aren't the same as facts. Instinct is not science. I'm like any other human being as fallible and as capable of being wrong. I only know what I believe.[17]

I'm not entirely sure what to make of "I only know what I believe" but, taking it in the broader context of the speech, including Blair's role as a politician and

a knowledge of the situation in Iraq, it appears to be an expression of passion, commitment, morality and fallibility. To interpret it as "for any x, if x is known by TB then x is also believed by TB but not vice versa because some of TB's beliefs may be false" would not be informative.[18]

As mentioned in Section 13.1, Stich and Ravenscroft distinguish between FP conceived of as 'the platitudes' and FP conceived of as an external systematisation of the platitudes. I suggest that the 'platitudes' are themselves the product of an external systematisation imposed on social life by certain philosophers. Some such systematisations may be very useful, providing ordered frameworks into which various phenomena can be conveniently grouped. But FP is not. What people actually say and do in social situations bears little resemblance to the use of 'belief' that characterises examples of so-called folk psychological platitudes. FP 'belief' groups together all manner of psychological phenomena and confuses an understanding of situational factors with the attribution of psychological states, resulting in examples of social explanation that are ambiguous, mistaken or irrelevant.

As Needham (1972, p. 125) puts it, the term 'belief' is a 'peg word' that is employed to refer to many quite different states. If this is accepted, there will be no set of systematic, law-like or regular connections to be discovered that relate beliefs to other states, such as desires and intentions, or to behaviours. As there is no kind of entity that is 'belief', there will be no set of informative generalisations applicable to all and only what we call 'beliefs'.

One response would be to restrict use of the term 'belief' to propositional attitudes. This would exclude a number of states that we ordinarily refer to as beliefs. For example, it seems highly unlikely that many religious beliefs are wholly articulable in terms of propositional attitudes. Consider the following passage:

> The extraordinary fact is that belief has survived such tests again and again – not because it comforts or explains but because believers cannot deny what has been shown or given to them. They have learned to see the world and life in the world as a freely given gift; they have learned to be open to a calling or invitation from outside their own resources, a calling to accept God's mercy for themselves and make it real for others; they have learned that there is some reality to which they can only relate in amazement and silence (Rowan Williams, Archbishop of Canterbury, Sunday Telegraph, 2nd January, 2005, referring to the 26th December 2004 tsunami).

'Belief', in this instance, is not of the form 'x believes that p'. It is instead a general orientation, a sense of conviction that permeates a life, shaping the manner in which one interprets things. It is arguable that many other 'beliefs' have a similar form and are not convincingly conveyed in terms of assent to one or more propositions. However, even if we restrict our account to those cases that can be conveyed in the form 'x believes that p', it is arguable that we still end up including too much. As noted earlier, believing that one's train is about to leave might take the form of a complex experiential state, whilst other 'beliefs' might involve assenting to the truth of a sentence or perhaps disinterestedly inferring that something is the case on the basis of an experience.

One way round this is to construe propositional attitudes more restrictively, in explicitly linguistic terms, leaving out other kinds of psychological states and an understanding of social situations. Although this might facilitate succinct use of the term 'propositional attitude', it would also result in a very restricted account that left out much that is important to everyday interpersonal understanding. Given that so many other factors contribute to an understanding of behaviour, the exercise of trying to predict or explain it on the basis of systematic connections between linguistic attitudes alone would, in most cases, be futile. For example, although emotions play a substantial role in most, if not all, of our activities, it is now generally acknowledged that emotions, whatever they turn out to be, are not propositional attitudes and are certainly not linguistic attitudes.[19] But if the emotions are removed from an account of FP, along with experience, habit, convictions and an understanding of equipment and social roles, what remains is an impoverished abstraction, which is ineffective when it comes to almost all cases of action interpretation. There are no systematic connections between intentional states, construed in this narrow, stripped down sense, and social behaviour, just as there are no systematic connections between red balls alone and the unfolding of a game of snooker. Hence FP 'belief', if construed in a general, encompassing way, is ambiguous and often misleading. However, if its scope is restricted, it is an impoverished abstraction from social life, which fails to illuminate the structure of social understanding.[20]

Rejecting the labels 'folk' and 'commonsense' and calling one's view something else will not circumvent these problems. So long as one retains the same explanandum, one inherits all the same problems. For example, Nichols and Stich (2003) prefer the term 'mindreading', given the worry that 'folk psychology' might incorporate "substantive theoretical commitments" (p. 2), in the form of a bias towards theory theory and against simulation. However, they assume the standard explanandum of how "the attribution of mental states to others" (p. 60) is achieved and use the term 'belief' throughout without any further qualification. They offer an intricate account of how mindreading is accomplished, incorporating a "motley array of mechanisms" (p. 212) but the first philosophical task is surely to settle on a tenable characterisation of kinds of social situation and of what we actually do in those situations. Without this, a theory of how we do it runs the risk of being not only false but also irrelevant. The cursory observations and arguments ventured here do not constitute a positive philosophical description of the structures of interpersonal and social understanding and interaction but they are, I hope, sufficient to reveal that the job of describing those structures is a difficult one, which is not done by FP.

ACKNOWLEDGEMENTS

I am grateful to Richard de Blacquiere-Clarkson, Daniel Hutto, Jonathan Lowe, Darrell Rowbottom and audiences at the universities of Central Florida, Hertfordshire and Hull for helpful comments on an earlier version of this chapter.

NOTES

[1] Although Churchland acknowledges that belief-desire psychology *is* the core of social life, he questions whether it *should be*, by arguing that (a) beliefs, desires and other mental states will have no place in a mature neuroscience and (b) commonsense psychology is not as successful as it is often made out to be.

[2] See Davies and Stone (1995a, b) and Carruthers and Smith (1996) for a variety of different simulation, theory and hybrid positions, most of which assume a common characterisation of 'folk psychology' as the attribution of intentional states in order to predict and explain behaviour. See Nichols and Stich (2003) for a recent hybrid theory, which again takes for granted the centrality of belief-desire psychology.

[3] Stawarska has since tried these questions out on a class of students in the form of a questionnaire. Responses to all questions were quite mixed, suggesting that no shared, intuitive view predominates (personal correspondence).

[4] Hobson (1993, p. 115; and also this volume) explicitly appeals to Strawson's view of persons in his discussion of the interactions between infants and care givers.

[5] There are, of course, considerable difficulties involved in determining whether a particular action does or does not accord with dualism.

[6] All comments are quoted verbatim. Students were informed that what they wrote might be published and that their names would be withheld.

[7] The term 'understanding' might have been employed as a synonym for 'explanation' but this is not something that one can assume, especially given the explicit contrast often drawn between a hermeneutic social science, which seeks 'understanding', and a naturalistic social science, whose task is 'explanation'.

[8] More recently, I distributed a questionnaire, which instructed students to look at eight different descriptions of interpersonal understanding and asked them to tick only the one that best characterised the way in which they understood others. Only one out of forty students ticked belief-desire psychology. See Ratcliffe (2007, Chap. 2) for further discussion.

[9] 'Folk psychology', construed as belief-desire psychology, did emerge from a context of explicit philosophical concerns (see introduction to this volume and Ratcliffe, 2007, Chap. 2). However, what I have not been able to find is an account of FP that starts off by carefully describing various social situations in order to show where and how FP operates. Instead, there is a tendency to presuppose FP on the basis of certain philosophical assumptions and then proceed to describe all manner of social situations in FP terms. That social situations can be described in terms of FP does not imply that they actually involve it, as will become clear in Sect. 13.5. What I am looking for here is an argument that arrives at FP through a study of social life, rather than an argument that arrives at FP through a debatable set of philosophical presuppositions and then proceeds to impose it on social life.

[10] See also Gallagher (2001), Hobson (2002), Hutto (2004), Ratcliffe (2005a, 2006) and Stawarska (this volume).

[11] For a more detailed discussion of this example, see Ratcliffe (2007, Chap. 4).

[12] See Hutto (this volume) for the claim that belief-desire psychology operates through interactive construction of shared narratives, rather than through theory or simulation.

[13] See Whiten and Byrne (1997) for some discussions of strategic interaction.

[14] Similar concerns about 'belief' are raised by Morton (2003, this volume).

[15] See Gurwitsch (1979) for a detailed phenomenological account of how situations, organised in terms of social roles and equipmental functions, enable meaningful social interaction and coordination. Gurwitsch takes social situations of this kind to be more fundamental constituents of social understanding than Schutz's 'Thou orientation'. See Gallagher (2005) for an account of how Gurwitsch's work can be employed to critique FP.

[16] Morton (2003, Chap. 3) observes that the folk seem to be very flexible about what is meant by 'belief'

[17] Thanks to Jonathan Lowe for this example.

[18] Similar concerns apply with respect to other so-called folk psychological concepts, such as 'desire'. If used in a general way to encompass any psychological factors that can motivate behaviour, it becomes a vague and uninformative placeholder. Consider the notion of commitment. In arguing that the unitary

conception of desire upon which rational decision theory rests is naïve and inaccurate, Sen (1977, p. 197) observes that counter-preferential choices are commonplace and stem from commitments to the maintenance of various public goods, such as parks, which structure one's activities even when they are countered by strong desires. According to Sen, these commitments are not simply conflicting desires but very different kinds of motivation.

[19] Most recent work on the emotions explicitly acknowledges that emotions are intentional, experiential states but not propositional attitudes. See, for example, the essays collected in Solomon (2004).

[20] It is arguable that an understanding of propositional attitudes, construed in a specific, linguistic sense, is at least useful when it comes to interpreting and explaining those actions that result from explicit reasoning processes, under the assumption that these processes are essentially linguistic (Hutto, this volume). However, I doubt that most instances of everyday reasoning can be understood wholly or largely in terms of the manipulation of linguistic structures. For instance, it is arguable that inarticulate 'feelings' of doubt and conviction and more general background feelings of what is and should be the case are integral to reasoning (Ratcliffe, 2005b). In most cases, our understanding of the reasons why people act is more inclusive, subtly distinguishing a range of factors that are either excluded from the orthodox view of FP or, alternatively, bundled together under the label 'belief' (Chap. 7 in Ratcliffe, 200).

REFERENCES

Bloom D (2004) Descartes' baby: how the science of child development explains what makes us human. Heinemann, London.

Bruner J (1990) Acts of meaning. Harvard University Press, Cambridge, MA.

Carruthers P, Smith PK (eds) (1996) Theories of theories of mind. Cambridge University Press, Cambridge.

Churchland PM (1998) Folk psychology. In: Churchland PM, Churchland PS (eds) On the contrary: critical essays 1987–1997. MIT Press, Cambridge, MA, pp 3–15.

Davies M, Stone T (eds) (1995a) Mental simulation: evaluations and applications. Blackwell, Oxford.

Davies M, Stone T (eds) (1995b) Folk psychology: the theory of mind debate. Blackwell, Oxford.

Dreyfus HL, Dreyfus SE (1999) The challenge of Merleau-Ponty's phenomenology of embodiment for cognitive science. In: Weiss G, Haber H (eds) Perspectives on embodiment: the intersections of nature and culture. Routledge, London, pp 103–120.

Frith U, Happé F (1999) Theory of mind and self-consciousness: what is it like to be autistic? Mind Lang 14:1–22.

Gallagher S (2001) The practice of mind: theory, simulation or primary interaction? J Conscious Stud 8(5/7):83–108.

Gallagher S (2005) Situational understanding: a Gurwitschian critique of theory of mind. In: Embree L (ed) Gurwitsch's relevancy for cognitive science. Kluwer, Dordrecht.

Gascoigne P (with H. Davies) (2004) Gazza: my story. Headline, London

Goffman E (1982) Interaction ritual: essays on face-to-face behaviour. Pantheon Books, New York.

Greenwood J (1999) Simulation, theory theory and cognitive penetration: no instance of the fingerpost. Mind Lang 14:32–56.

Gurwitsch A (1979) Human encounters in the social world. Transl. by Kersten F. Duquesne University Press, Pittsburgh, PA.

Hobson P (1993) Autism and the development of mind. Erlbaum, Hove.

Hobson P (2002) The cradle of thought. Macmillan, London

Hutto D (2004) The limits of spectatorial folk psychology. Mind Lang 19:548–573.

Langdon R, Davies M, Coltheart M (2002) Understanding minds and understanding communicated meanings in schizophrenia. Mind Lang 17:68–104.

Libet B (2004) Mind time: the temporal factor in consciousness. Harvard University Press, Cambridge, MA.

MacDonald C (2002) Theories of mind and the 'commonsense view'. Mind Lang 17:467–488.

Morton A (2003) The importance of being understood: folk psychology as ethics. Routledge, London.
Needham R (1972) Belief, language, and experience. Blackwell, Oxford.
Nichols S, Stich SP (2003) Mindreading: an integrated account of pretence, self-awareness, and understanding other minds. Clarendon Press, Oxford.
Ratcliffe M (2005a) Folk psychology and the biological basis of intersubjectivity. In: O'Hear A (ed) Biology and life: Royal Institute of Philosophy supplement 55. Cambridge University Press, Cambridge, pp 211–233.
Ratcliffe M (2005b) The feeling of being. J Conscious Stud 12(8–10):43–60.
Ratcliffe M (2006) Folk psychology is not folk psychology. Phenomenol Cogn Sci 5:31–52.
Ratcliffe M (2007) Rethinking commonsense psychology. Palgrave Macmillan, Basingstoke.
Scholl BJ, Leslie AM (1999) Modularity, development and 'theory of mind'. Mind Lang 14:131–153.
Schutz A (1967). The phenomenology of the social world. Transl. by Walsh G, Lehnert F. Northwestern University Press, Evanston.
Sen AK (1977) Rational fools: a critique of the behavioral foundations of economic theory. Philos Public Affairs 6:317–344.
Solomon RC (ed) (2004) Thinking about feeling: contemporary philosophers on emotions. Oxford University Press, Oxford.
Stich SP, Ravenscroft I (1996) What *is* folk psychology? In: Stich's Deconstructing the mind. Oxford University Press, Oxford, pp 115–135.
Strawson PF (1959/1990) Individuals. Routledge, London.
Wellman HM (1990) The child's theory of mind. MIT Press, Cambridge, MA.
Whiten A, Byrne RW (eds) (1997) Machiavellian intelligence II: extensions and evaluations. Cambridge University Press, Cambridge.
Wimmer H, Perner J (1983) Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13:103–128.

NAME INDEX

SUBJECT INDEX