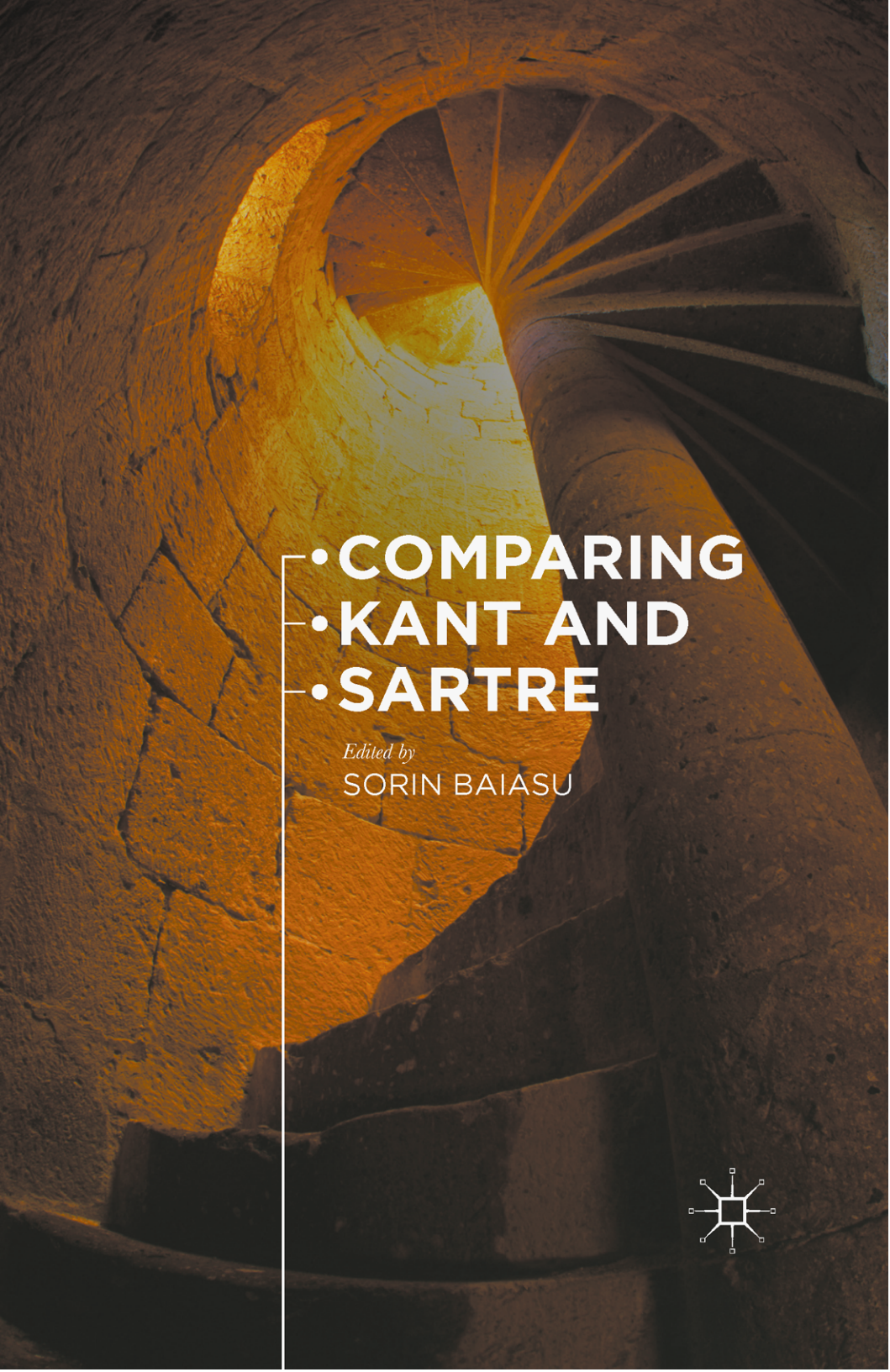
COMPARING
KANT AND
SARTRE
Edited by
SORIN BAIASU

Comparing Kant and Sartre

Also by Sorin Baiasu

KANT AND SARTRE: Re-discovering Critical Ethics
POLITICS AND METAPHYSICS IN KANT (*co-edited with Sami Pihlström and Howard Williams*)
KANT ON PRACTICAL JUSTIFICATION: Interpretive Essays (*co-edited with Mark Timmons*)

Forthcoming Title

SINCERITY IN POLITICS AND INTERNATIONAL RELATIONS (*co-edited with Sylvie Loriaux*)
THE KANTIAN MIND (*co-edited with Mark Timmons*)

Comparing Kant and Sartre

Edited by

Sorin Baiasu
Keele University, UK

palgrave
macmillan

Softcover reprint of the hardcover 1st edition 2016 978-1-137-45452-2

First published 2016 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

ISBN 978-1-349-55673-1 ISBN 978-1-137-45453-9 (eBook)
DOI 10.1057/9781137454539

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Baiasu, Sorin.
Comparing Kant and Sartre / Sorin Baiasu, University of Keele, UK.
pages cm

1. Kant, Immanuel, 1724–1804. 2. Sartre, Jean-Paul, 1905–1980. I. Title.

B2798.B25 2015
193—dc23 2015023919

In Memory

Doina Sabina Baiasu
1938–2015

Exemplary Mother, Selfless Friend, Inspiring Mentor

Contents

Part IV Metaphilosophy

List of Contributors

Justin Alam, English for Academic Purposes Tutor, Centre for English Language and Foundation Studies (CELFS), University of Bristol, UK

Richard E. Aquila, Professor, Associate Department Head and Director of Graduate Studies, Department of Philosophy, The University of Tennessee, USA

Sorin Baiasu, Reader, Philosophy Programme Director and Chair of the Keele-Oxford-St Andrews Research Centre for Kantian Studies (KOSAK), Keele University, UK

Jochen Bojanowski, Faculty Member, Department of Philosophy, University of Illinois at Urbana-Champaign, USA

Michelle R. Darnell, Lecturer, Department of Management, University of Florida, USA

Thomas R. Flynn, Samuel Candler Dobbs Professor, Department of Philosophy, Emory College of Arts and Sciences, USA

Jonathan Head, Graduate Teaching Assistant, Department of Politics, International Relations and Philosophy, Keele University, UK

Daniel Herbert, Department of Philosophy, The University of Sheffield, UK

Katherine Morris, Supernumerary Fellow in Philosophy, Mansfield College, University of Oxford, UK

Christian Onof, Honorary Research Fellow, Department of Philosophy, Birkbeck University of London, UK

Peter Poellner, Professor, Department of Philosophy, University of Warwick, UK

Leslie Stevenson, Honorary Reader, Department of Philosophy, University of St Andrews, UK

Anna Tomaszewska, Post-doctoral Researcher, Department of Philosophy, Jagiellonian University, Poland

Alberto Vanzo, Research Fellow, Department of Philosophy, University of Warwick, UK

Part I
Introduction

1

Kant and Sartre: Existentialism and Critical Philosophy

Jonathan Head, Anna Tomaszewska, Jochen Bojanowski, Alberto Vanzo and Sorin Baiasu

1 Kant and Sartre so far

Kant and Sartre are two of the most significant figures in modern philosophy, and yet there has, until very recently, been little comparative research undertaken on them. Despite dealing with many shared philosophical issues, they have traditionally been taken to be too opposed to each other to render any search for possible parallels between their works a useful enterprise. Indeed, Sartre is often taken to be one of Kant's most vocal critics in the literature, and as rather indebted to other major figures, such as Husserl and Heidegger. As a consequence, often, where comparative analysis has been done upon Kant and Sartre, the emphasis has been on their differences, rather than on their similarities. However, as recent research has begun to show, the story is not that straightforward and there is much to be explored with regard to parallels between Kant and Sartre. Baiasu (2003) has characterized Sartre's relation to Kant as one of an "anxiety of influence" – Sartre desires to explicitly distance himself from Kant, but this obscures some deeper underlying parallels between them.[1] Such parallels can form a foundation for productive

Acknowledgement SB: Work on this chapter was carried out while an Honorary Guest Research Professor at the University of Vienna, as part of the ERC Advanced Research Project "Distortions of Normativity". I am grateful to the project's PI for making this possible.

[1] Howells (1988) originally made this point with regard to their respective positions on ethics, which Baiasu (2011) later expanded to cover wider aspects of their philosophies.

dialogue, more widely, between the schools of Kantian "Critical philosophy" and existentialism.[2]

Recent research has demonstrated the possibility of such a dialogue between the philosophies of Kant and Sartre. A natural starting point for comparative analysis of both philosophers is that of their ethical theories, which has sparked differences of opinion among scholars. Linsenbard (2007), for example, has argued that Sartre's use of Kantian notions (such as a principle of universalizability) masks more fundamental differences between the two that place them far apart:

> [I]t would be a mistake, I think, to interpret Sartre's views on morality as 'Kantian' or as even marginally endorsing Kant's views. Indeed, Sartre's continuing preoccupation (one might even say 'obsession') with Kant suggests … a path he did *not* wish to take with respect to the most promising moral terrain. (2007: 65)

Due to having radically different ethical theories, despite much talk of Kant and use of familiarly Kantian language, Sartre "cannot … be interpreted as invoking Kant's meaning" (2007: 80). Painter (1999), on the other hand, finds deep similarities between Kant and Sartre on the questions of ethics, identifying a shared inheritance in the tradition of Protestant ethics. Describing Sartre's relationship with Kant as "flirtatious", he writes that

> [b]oth find a common ground in a fundamental aspect of the Protestant ethic, characterized by Lutheranism and Calvinism, wherein the everyday takes on great moral significance, and great deeds, or high moral principles that direct actions based on the actualization of virtuous ends become hubristic, impious and immoral … how we approach the simply given in life, the concrete everyday situation, has far more moral significance than any moral principle whose content defines what is right or what is wrong. (1999: 211)

Given this shared inheritance, we can see Sartre's use of Kant's language and various concepts as an opportunity to illuminate ethical insights from the Protestant ethical tradition, alongside his own idiosyncratic

[2] An example of such a dialogue can be seen in the work of Christine Korsgaard, who explicitly acknowledges her indebtedness to both Kant and Sartre. Whilst her work is more widely recognized as Kantian in spirit, nevertheless it also includes strong existentialist aspects.

developments, and Sartre himself as unable to escape from shared parallels with Kant on the question of ethics: "Sartre's critique of Kant's ethics, and his attempt to develop his own, burns down, like a crucible, the essence of both approaches: a secular Protestantism" (1999: 217).

As part of the project of a comparative analysis on Kant's and Sartre's ethical works, Sweeney (1985) has also noted that, in his short story "The Wall", Sartre uses examples similar to those of Kant's famous essay "On a Supposed Right to Lie From Philanthropy", which illuminate the ethical theses that he is attempting to illustrate through the narrative.[3] We see, in this story, Sartre potentially using Kantian resources to argue against Husserl; as Sweeney writes, "Sartre seeks to argue against Husserl by presenting through his use of Kant's example a counter-example to Husserl's view" (1985: 15). Though, of course, this in itself does not show that Sartre is adopting a Kantian ethical theory, it does illustrate at least that he was aware of the philosophical resources made available to him by Kant for use in describing and elaborating his own ethical theory.

Lieberman (1997) has also added to this literature by comparing Kant and Sartre's accounts of freedom, particularly taking into account the impact that radical evil has upon freedom in *Religion within the Bounds of Mere Reason*. An interesting parallel can perhaps be seen between the choice of a fundamental ethical disposition as an original act of will (which, as far as Kant is concerned, makes all human beings radically evil) and Sartre's notion of 'choosing oneself' – both are choices which are independent of the individual's environment, are "outside of time", and serve as an intelligible ground for individual choices, seeing them as part of a "total choice" (1997: 210–12).

However, Lieberman is keen to note that the parallels between Kant and Sartre on the notion of a kind of fundamental ethical choice only go so far; ultimately they differ insofar as "first, Sartre lacks the world view that accepts common (and perhaps unquestionable) knowledge of the moral law; and second, Sartre lacks the theoretical orientation in which an a priori awareness of the moral law is possible – in which a fact about our essence as rational beings precedes, or is at least independent of, our existence – thereby reversing the existential canon that existence precedes essence" (1997: 215). Nevertheless, he speaks of the comparative analysis of Kant and Sartre on this issue as "fruitful" in "revealing in the residue of analysis an historical trace that connects Kant to Sartre in some aspects of their thought", as well as "[bringing]

[3] Though, Sweeney does allow that Sartre could have been inspired to use this example by Plato or Victor Hugo (1985: 16 n7).

to light fundamental problems within their theories and [suggesting] avenues of interpretation and possible solutions" (1997: 216).

Other recent research has focused on issues in the theoretical philosophies of Kant and Sartre. Deep parallels and dissimilarities have both been noted. As an example, Jopling (1986) has discussed their accounts of self-knowledge (a topic that will recur in this volume). A deep affinity can be found in the sense in which the attempt to gain self-knowledge, for both philosophers, is a very difficult endeavour indeed:

> Kant and Sartre, I believe, are calling attention to the existence of a blind spot which unavoidably insinuates itself into all our attempts to know ourselves. The activities necessary for self-knowledge … are always one logical step behind themselves, and are blind to the very agency constitutive of and contemporaneous with them. We are unable to know ourselves in the very act of knowing … We know ourselves through the categories, or through the 'Other' – and not as absolutely proximate and self-present. (1986: 74)

Sartre follows Kant's approach in seeking "to correct the strong tendency towards reification and substantialization which infects … both philosophical and pre-philosophical self-knowing activities" (1986: 75) and holding a "radicalized version of the concept of constituting activity" (1986: 73) that ultimately denies straightforward knowledge of the self. More recently, Darnell (2005 – a contributor to this volume) has also published a monograph on the notion of self in Kant and Sartre, noting the complex relations between the two on this topic, and that Sartre's misreading of Kant may have led him to distance himself from Kantian thought more than he needed to; for example, "he most likely fell victim to Kant's characterization of the I of apperception as not only a unity, but also as a ground of identity" (2005: 27). Also as part of this body of research on theoretical philosophy, Gardner has recently considered the extent to which Sartre can be labelled a "transcendental philosopher". He paints a complex picture of the position of Sartre in the post-Kantian tradition, but nevertheless argues that in a substantial sense Sartre can be seen as following a Kantian line in some of his theoretical thinking. As an example, Gardner points to Sartre's "anti-naturalist strategy" as being "at least in substantial part, transcendental" (2011: 54) due to the use of recognizably transcendental argumentation.

Of course, there is much more literature now available as part of the growing body of research on Kant and Sartre – these have been merely examples to give at least a limited sense of the wealth of research

opportunities available in comparing these two philosophers. This volume of original essays is intended to be a significant addition to the growing body of comparative research on Kant and Sartre, encompassing in an unprecedented manner a number of original papers that embrace many philosophical topics of interest shared between the two thinkers. Many of the papers stem from a conference on Kant and Sartre held at Keele University in November 2012. The volume, split into three major parts, addresses issues in metaphysics, metaethics and metaphilosophy. Philosophical notions central to both Kant and Sartre, including autonomy, happiness, self-consciousness, self-knowledge, evil, temporality and the imagination are explored in great detail to give us a clearer picture of the theoretical and practical philosophies of both thinkers. In addition to giving us new insights, the papers also leave many unanswered questions and thus give us promising prospects for future comparative research on Kant and Sartre. The rest of this Introduction will discuss some of the key points of the papers in the order in which they appear in the volume, as well as the issues and difficulties they raise for future research.

2 Comparing Kant and Sartre

2.1 Metaphysics

The volume proper begins with a contribution by Sorin Baiasu, who considers two objections generated by his claim that Kant's transcendental unity of apperception is deeply similar to Sartre's (self-) consciousness or pre-reflective consciousness of self. Both objections are prompted by Baiasu's claim that Sartrean pre-reflective consciousness of self and Kantian transcendental unity of apperception play the role of a weak epistemological condition of experience. The first objection indirectly challenges the weakness of such a condition, whereas the second disputes its epistemological nature. The first objection is said to have implications for the current debate concerning non-conceptual content, whereas the second is regarded as linked to traditional debates concerning the kind of idealism that Critical philosophy and phenomenology offer. After a brief discussion concerning some methodological problems for comparative philosophical studies, the chapter answers the two objections and examines their implications.

However, the weak sense of personality important to Baiasu's argument here, that is, the epistemological condition offered by the transcendental unity of apperception, which is necessary in a practical,

moral sense, raises questions. We can reflect, for example, upon a shift in Kant's thought between the first and second editions of the first *Critique*; in the former, Kant denies knowledge of substance in a thick sense, although he seems to allow that we may be substances in a thin, logical sense, whereas in the latter, he denies even this sort of thin knowledge of ourselves as substances. It would perhaps be an interesting line of future research to evaluate how a comparison with Sartre might shift if we concentrate on the differing first and second editions of the first *Critique*. Furthermore, questions can be raised as to whether there is in fact a notion of a weak sense of personality in the A edition.[4] Such uncertainty regarding Kant's position further complicates the issue of how we compare his commitments in the first *Critique* with Sartre's pre-reflective consciousness of self.

Continuing on the theme of the conditions of our experience, Daniel Herbert's chapter focuses on the topic of temporality. He argues that a fundamental misunderstanding of transcendental idealism as involving an ontological commitment to a supersensible reality leads Sartre to make unfair criticisms of Kant's treatment of temporality. If we opt for an Allison-style 'methodological' or 'two-aspects' reading of transcendental idealism (where we refrain from stating that the objects of experience are ideal, even if space and time are), we can see that Sartre's criticisms regarding the linking of temporality to the perspective of a transcendental subject are perhaps ill-founded. However, it is not clear that traditional worries concerning the Kantian use of the thing-in-itself are avoided under the two-aspects model, for if the forms of intuition are ideal, and intuition is supposed to provide evidence for the reality of sensible objects, then we have the worry that sensible intuitions themselves and the 'reality' of the objects are ideal too.

Nevertheless, Herbert further reflects upon where the two philosophers diverge with regard to temporality; whereas Kant's account is more impersonal, Sartre desires to ground his understanding of temporality in everyday experience, in particular through our capacity for spontaneity, and not making a distinction between a transcendental 'extra-mundane' subjectivity and its empirical counterpart. Kant's overemphasis on the mathematical sciences leads him, in Sartre's view, to posit an unacceptably strong distinction between the empirical and the noumenal. Whilst Sartre recognizes the relevance of temporality in all domains of human activity, Kant seems to limit this to the domain of science alone.

---

[4] Ameriks has argued that some interpreters are mistaken in ascribing personality to the 'I' (2000[1982]: esp. ch. 4).

The discussion opens up further directions for research; for instance, we can examine whether Sartre's interpretation of Kant on temporality, the foundation on which he forms his objections, is accurate. What seems objectionable is that Sartre sees Kant's theory of time as relevant to a theory of science but not at the same time to a theory of mind and human cognition. Yet, in fact, Kant's transcendental philosophy and his account of time can be read as (part of) a theory of the conditions of the possibility of science (e.g. De Vleeschauwer 1962) or a theory of mind and cognition (e.g. Kitcher 1990) or a theory of experience (e.g. Aquila 1983). The point is that Kant's account of time has been read in more "phenomenological" ways too, and that for comparing him with Sartre this could prove fruitful.

With Thomas R. Flynn's contribution, we turn from temporality to the imaginary. His chapter reflects upon Kant's influence on Sartre's psychology, with a particular emphasis on the imagination. Flynn works through a number of Sartre's works, noting potentially illuminating parallels between the two philosophers. As an example, the young Sartre seemed to have been enamoured with the role that the imagination has to play in the *Critique of Judgement*, along with Kant's use of symbolic schematism. Flynn also notes points where Sartre seems to have been spurred by Kant to develop certain aspects of his own philosophy, such as the notion of an 'egoless' consciousness, the placing of the imaging consciousness at the very centre of his philosophical psychology, and the appeal to the 'as if' in expanding our imaginary reflections upon philosophical issues (in a parallel with Kantian regulative ideas). The chapter concludes with reflections upon the parallels and tensions between Sartre's later ethics and Kantian moral theory.

There are open questions here regarding the distance between Kant and Sartre, and indeed whether the latter sees himself as attacking the former. Flynn argues that the 'egoless' consciousness forms part of Sartre's attack on Kant, alongside intentionality and a realist epistemology, posed against Kant's constitutive character of consciousness. Is Sartre simply interested in different questions than Kant, and can we construe Kant along Sartrean lines, with himself adopting intentionality and a realist epistemology? Is the Kantian constitutive character of consciousness so unamenable to Sartre?

The part of the volume on metaphysics concludes with a chapter by Christian Onof on the key metaphysical notions of freedom and the self in Kant and Sartre. More specifically, Onof attempts to use philosophical resources from Sartre to aid Kant with his problem of reconciling

transcendental freedom with causal determinism in the Third Antinomy. This account appears to leave us with something of a dilemma with regards to the possibility of evil – if we assume that to be free means to be moral, then one can only be evil if one is not free, and hence not responsible for their evil actions. (Perhaps we could attempt to resolve the dilemma by allowing for different types of freedom in addition to autonomy, which is a presupposition for being free). We can fill out our understanding of Kant's views on these points by noting his distinction between *Wille* (practical reason under the moral law) and *Willkür* (the legislative – transcendentally free – power of choice). However, we may still wonder how *Willkür* chooses noumenally, particularly when it does not legislate in accordance with the moral law.

Onof suggests that Sartre's ontology of the "For-Itself" and the "In-Itself" can aid Kant here in placing spontaneity outside of being, and considers whether this could aid with the potential difficulty that God's creative act could predetermine our actions. However, Sartre's account leads to an emphasis on consciousness as the source of negation in a way that the Kantian would resist. Instead, Onof concludes with the suggestion of a modal realist interpretation of transcendental idealism in order to maintain normativity as being connected to what is possible. He also argues that this position is compatible with Kantian moral theory.

The chapter raises many questions, as well as tapping into a wider debate (seen in a number of contributions to this volume) regarding how dissimilar Sartre's "realist" ontology is from Kantian transcendental idealism. As an example, we could consider what "outside" means with regard to the For-Itself's lying outside of the fullness of the being of the In-Itself. A natural reading would be to take 'outside' as denoting ontological distinctness, but what kind of distinctness this could be, given that 'outside' of being is outside of ontology itself, is a further question to be considered. In relation to this, how are we to understand "nothing" in a Sartrean context? Is it absolute, in denoting there not being anything, or merely relative, as a positive something with a specific role to play in Sartre's philosophy? Indeed, how are we to understand more generally the status of such claims regarding what is beyond being? We may also raise questions about Kant, for example, whether the notions of substance and causation (both a priori concepts of the understanding) should be modified when faced with the difficulties of accounting for freedom, instead of merely shifting to a different notion of causation.

2.2 Metaethics

In the following part of the volume, our attention moves from metaphysics to metaethics, beginning with Leslie Stevenson's essay on self-knowledge and its relation to freedom in Kant and Sartre. Stevenson begins by reflecting upon the Sartrean "pre-reflective cogito", which bears more than a passing resemblance to the Kantian 'I think' of the transcendental unity of apperception. Using a contrast with animal mentality, he argues that both Kant and Sartre seem to have a sense that human beings can have conceptualized perceptions that are unavailable to other animals, such that we can become "positionally" aware of ourselves in relation to our environment and explicitly aware of other facts about ourselves. In this regard, Stevenson utilizes the pure/impure reflection distinction in Sartre to suggest that we could understand such self-awareness as a kind of "purifying" reflection upon our own fundamental purposes. Finally, the chapter concludes by arguing for a deep parallel between Kant and Sartre on the question of self-knowledge in relation to freedom, in that, for both, self-knowledge (in whatever way you wish to construe it) and project-setting can act in tandem.

Further reflections upon Sartre's account of self-knowledge could focus on the claim, in *The Transcendence of the Ego*, that non-reflective consciousness is an impersonal, transcendental field of consciousness. If such an impersonal consciousness is the basis for reflective as well as positional consciousness, then can reflective consciousness be said to be personal? It becomes difficult to see how self-knowledge comes from reflective consciousness, given that, according to the early Sartre, it provides an ego that is part of the world and thus not part of a person's consciousness. The approach of comparing human and animal consciousness in Kant and Sartre is also promising, though it poses further questions. One such question revolves around how we should distinguish animal and human consciousness in the context of Sartre's various commitments regarding consciousness. We may think that animals should not be attributed pre-reflective cogito because the human pre-reflective consciousness of self is what allows us to reflect upon our actions in a way that other animals do not. However, is this enough to deny pre-reflective cogito to animal consciousness, particularly if we may want to grant them pre-reflective positional consciousness, which (according to Sartre) rests upon a pre-reflective consciousness of self? Drawing the lines between human and animal consciousness in this way, within the confines of Sartre's philosophy, is certainly a research direction worth pursuing.

Peter Poellner's chapter focuses upon autonomy in Sartre's philosophy, with a view to drawing lessons for contemporary debates surrounding practical rationality. To begin with, he explores the Sartrean view of seeing autonomy as tied to a consciousness being both self-determining and sensitive to reasons. Poellner delineates a number of aspects of Sartre's account of freedom, and argues that these do not have direct metaphysical import, but rather refer to phenomenological facts, to a "practical reality of action". Under this view, autonomy can be seen as the very foundation of an "ethics of freedom", and there lies a possible parallel with Kant. He further reflects upon Sartre's "completion thesis". Sartre appears to have taken this thesis to have had a certain amount of ontological import, stating that, in pursuing ends, consciousness experiences a lack and feels 'incomplete' in a sense. Thus, a corollary of action is a desire on the part of consciousness to complete itself, and overcome the lack that it has previously felt. Poellner, having undertaken a certain amount of "rational reconstruction", evaluates this "quasi-Sartrean" view and finds it somewhat wanting in its obscurity. Nevertheless, he thinks we can formulate a sufficiently filled-out view such that we can evaluate it in comparison with Kantian ethical theory. In particular, with the notion of a consciousness' having value insofar as it aims at unqualifiedly valuable ends, a value that potentially encompasses all conscious beings, we may be reminded of the Kantian quest for the *summum bonum*, or the "highest good".

Poellner's chapter suggests a further avenue of research when he distinguishes between fundamentally value-centred and reasons-centred views, with Sartre falling under the former umbrella and Kant under the latter. Does this distinction show a fundamental discontinuity between the two philosophers? It is especially noteworthy that Kant's ethical theory is, in a sense, value-centred, taking good and evil not as values to be derived from normatively neutral features, but in fact as a priori ideas of reason, as we can see from the second *Critique*.

Justin Alam's contribution focuses upon two key concepts in the metaethics of Kant and Sartre: respectively, radical evil and bad faith. In *Religion within the Boundaries of Mere Reason*, Kant introduces the notion of a "supreme maxim", chosen as a principle for deciding whether to prioritize morally good or amoral incentives in selecting first-order maxims of action. It is in the selection of an evil supreme maxim that radical evil consists. Alam examines how we can understand the selection of an evil supreme maxim given that we have an overriding reason, following from our autonomy, to choose a good supreme maxim. In order to unpick this difficulty, he draws upon the possibility of

self-deception playing a part in the selection of a supreme maxim. However, self-deception itself brings problems with it, such as how we can at the same time both know and be ignorant about a given deception.

It is at this point that Sartrean bad faith comes into the picture. As a way of undermining the difficulties surrounding the notion of self-deception, it provides the possibility of a situation in which an individual does not simultaneously believe one thing and its opposite, but rather distracts themselves from the truth by misinterpreting evidence for it. On this basis, Alam assesses the possibility of a Kantian use of this sort of strategy to make sense of the choice of the "supreme maxim". The idea of a will choosing a fundamental disposition on the basis of reasons, for example, may still retain an air of mystery (indeed, it may seem that Kant accepted that evil actions are ultimately unintelligible).

It is suggested that a kind of "false freedom" is the promise upon which we decide to prioritize self-love, but the possibility remains that the choice of freedom is more fundamental than the choice to prioritize self-love on Kant's model; or perhaps we are putting things the wrong way around, for freedom has a secondary value to the good: freedom makes the universal agreement of all rational beings possible, but it is the good that is the object of pure practical cognition. Self-deception itself, though, may also remain a wholly intractable problem. There is still a level of pretence involved in bad faith, contrasted with a sincerity condition required for true belief. Thus, we have a notion of an individual non-sincerely pretending something to be the case, which ultimately may remain just as mysterious as the idea of an individual simultaneously believing in one thing and in its opposite. All these questions foreshadow interesting avenues for future research.

The metaethical part of the volume concludes with Michelle Darnell's analysis of the role of happiness in Kant and Sartre, which she argues could play a larger part in the philosophy of both thinkers than is usually supposed. According to Darnell, Sartre criticizes Kantian ethics on the basis that it is not a positive ethics; being grounded in a noumenal realm, it is too far removed from the level of concrete events. However, Kant can offer a positive ethics drawing upon the notion of happiness, and the importance of interdependence of persons in creating value in the world, with the end-result of *summum bonum*, where happiness is established in proportion to virtue. Once we understand this positive aspect of Kant's ethics, which involves a teleological dimension of the unification of all moral ends, we can see that Sartre's criticisms are largely unfounded. In stressing this positive aspect of Kant's ethics, though, have we moved away from the key Kantian emphasis on autonomy?

In response to Sartre's objections, it seems promising to make the highest good into a social good upon which we ground principles of action. There is a danger, though, that if we make happiness in proportion to virtue a moral principle, we may be placing a fundamentally heteronomous principle at the heart of Kant's ethics, and if so, sacrificing a part of his ethics that makes it identifiably *Kantian*. Do we, in replying to Sartre's objections, grant him too much? Further, Darnell goes on to consider a possible role for happiness in Sartre's ethics. At first, it seems that such a role may be entirely negative as an expression of bad faith, marking out a means by which we seek to deny our own existence as for-itself. The beginnings of a possible positive role for happiness comes through a link to authenticity, in which our understanding of happiness can undergo a revolution, becoming a feeling of joy in response to the authentic person's free creation of a meaningful world. Thus, perhaps a deeper relation, grounded in a fundamentally optimistic attitude, can be found between Kant and Sartre, though questions regarding the kind of happiness that both philosophers have in mind remain. Sartre's view of happiness as consisting in an authentic life seems more akin to a Stoic model than the view held by Kant.

2.3 Metaphilosophy

The final part of the volume deals with metaphilosophical issues, beginning with a chapter by Katherine Morris on the possibility of reading Sartre as a "philosophical therapist", along the lines of Wittgenstein. The chapter, among other things, focuses on Sartre's practice of giving descriptions of everyday experiences, and using such descriptions as a springboard for reflections upon the phenomenology of human reality. Adopting a therapeutic approach involves moving from descriptions of experience to phenomenological claims in a way that breaks down resistance to such descriptions due to bad faith. Such an approach stands in contrast to what she calls a "transcendental reading", in which transcendental arguments are used in order to demonstrate that phenomenological claims are conditions for the possibility of certain kinds of experience. Through the examination of examples given by Sartre concerning the phenomenology of everyday experiences, Morris argues that the descriptions involved are not incontestable such that they can stand as a foundation for a transcendental argument; rather, Sartre can be read as concerned with bad faith as willing to misconstrue phenomenological ontology through resisting certain descriptions of experience. In this regard, we can draw upon Wittgenstein, who attempts to use philosophy to dispel intellectual prejudice, as difficult as that task may be.

Morris ends her chapter by answering possible objections to a therapeutic reading of Sartre. One such objection is that Sartre, on this reading, has left behind philosophical argumentation entirely, and another connected objection is that in this way Sartre has decisively left epistemological issues aside. Such objections can be met, Morris argues, through reflections upon Sartre's specific target on individuals with bad faith. This discussion, in turn, certainly invites further questions regarding Kant, for example, whether we could attempt to interpret Kant along therapeutic lines.[5] As a starting point for such reflection, Morris notes the impact the therapeutic reading has on the distinction between appearance and reality. The transcendental reading assumes that the distinction in question maps onto a gap that needs to be bridged, whereas the therapeutic reading paints Sartre as undermining the "scandal of philosophy" that is the difficulties surrounding an inference from appearance to reality. If a therapeutic reading of Kant proves successful, would this affect our evaluation of the relation between Kant and Sartre?

The volume concludes with a consideration of transcendental idealism in Kant and Sartre by Richard Aquila. He argues that both philosophers espouse a form of transcendental idealism called "transcendental phenomenalism", in which a judgement of an appearance as "real" involves both, from one point of view, affirming the appearance as phenomenon and, from another, affirming that it is an appearance within an infinite series of appearances of the phenomenon involved. Such a similarity is possible, if one construes transcendental idealism more generally as stating of a "real" phenomenon that a judgement is being made of a consciously-available object whose ontological status is left open. An empirical judgement requires an appeal to being-in-itself in addition to an appeal to infinity with regard to those species of appearances to which the objects in question are reducible. The appeal to being-in-itself, however, is not to be considered as regarding said objects as things-in-themselves from a transcendental perspective. It is through such parallels that Kant and Sartre can be seen as holding similar positions.

Nevertheless, differences between Kant and Sartre remain, for example, surrounding the distinction between the Kantian "thing-in-itself" and the Sartrean "being-in-itself". Aquila's chapter invites deeper metaphilosophical questions regarding the framework within which we compare Kant and Sartre. How are we to deal with technical terms like 'phenomena'

[5] This is a line taken by Graham Bird (for instance, 2006).

and 'appearance' within the framework of different philosophies? Taking appearance and phenomena as our example, not only might Kant and Sartre mean different things by these terms, but also the relationship between the two terms may differ, making any fruitful comparison between the two on these issues very difficult indeed. Further research of such issues is invited here.

3 Conclusion

So, to conclude, what questions and possible future avenues for research do these chapters raise? To begin with, we can consider wider issues surrounding the kind of comparative analysis undertaken here, particularly when one philosopher (in this case, Sartre) is consciously reacting to another (Kant). In such a situation, not only do we have to attend to varying competing interpretations of the positive philosophies of both figures, but we also have to be careful in placing the later philosopher's interpretation of the earlier philosopher among these various options. In the case of Kant and Sartre, a comparative analysis of these two philosophers may be very illuminating, and in particular may help to clarify various aspects of Sartre's philosophy and how he sees his own position in the history of philosophy. However, we may be misled if we are not clear on the question of which interpretation of Kant he is reacting to, a factor which would be crucial for any successful comparative analysis. Indeed, does it matter if Sartre is unfair to Kant, or if he is a perceptive, sensitive reader of him?

In addition to such an endeavour, we may still desire to compare Kant and Sartre on their own terms, regardless of how the latter interpreted the former, but we must be clear that this is an entirely separate issue. Further reflection could take place on which approach we find the most useful for our philosophical research – there may be a substantive difference between the two, they may complement each other very well, or it may just be a matter of taste, with differing aims for research in the history of philosophy. In addition to this, a number of chapters in this volume have attempted reconstructive work on the philosophies of Kant and Sartre in light of the contrasts and parallels between the two. However, this raises the question of how far we should go with such reconstructive work, and whether there is a point at which we lose something essential from the philosophy that is being reconstructed. There is perhaps a point in which we should leave a philosophy as it is, despite the difficulties it faces, in order to preserve unique insights that we may wish to plunder for our own philosophical needs.

Leaving methodological issues aside, these chapters have also raised a number of issues pertaining specifically to comparative research on Kant and Sartre. One general question we may wonder about is the extent to which Kant can be viewed as a kind of 'proto-existentialist', or if that is too much, the extent to which the existentialists draw crucial insights from specifically Kant's thought (as opposed to Kantian thought more generally). In addition, there are important metaphilosophical questions to answer regarding how Kant and Sartre both regard the aims and methods of their philosophies. To take an example, Kant's project in the Critical period is very carefully constructed (and rather idiosyncratic) to respond to specific issues in philosophy at that time, a situation that had certainly moved on by the early 20th century. Is there a sufficient amount of crossover between Kant and Sartre on the question of metaphilosophy to substantiate substantive conclusions through comparative analysis? Or, are there simply fundamental discontinuities between these two philosophers on these key issues?

Such worries also impact upon any attempt to focus on specific topics within their philosophies – for example, if one philosopher's treatment of freedom has very different aims and methods than another philosopher's account, then it will be very difficult to make secure, substantive comparative points on that topic. Nevertheless, the chapters in this volume show that, despite difficulties, much can be done in comparative research of Kant and Sartre. The two philosophers do have a great deal to say to each other, as well as to us. In conversation, they not only illuminate aspects of the philosophy of their interlocutor but also parts of their own. Doubtless, this conversation will continue in future research with a great amount of success.

References

Ameriks, K. (2000[1982]) *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. New edn. Oxford: Clarendon Press.

Aquila, R. (1983) *Representational Mind: A Study of Kant's Theory of Knowledge*. Bloomington: Indiana University Press.

Baiasu, S. (2003) "Sartre's Search for an Ethics and Kant's Moral Theory", *Sartre Studies International,* 9(1): 21–53.

—— (2011) *Kant and Sartre: Re-Discovering Critical Ethics*. London: Palgrave Macmillan.

Bird, G. (2006) *The Revolutionary Kant: A Commentary on the* Critique of Pure Reason. Chicago, IL: Open Court.

Darnell, M. (2005) *Self in the Theoretical Writings of Sartre and Kant: A Revisionist Study*. Lewiston: The Edwin Mellen Press.

De Vleeschauwer, H.-J. (1962) *The Development of Kantian Thought: The History of a Doctrine*. Tr. A. R. C. Duncan. London: T. Nelson.
Gardner, S. (2011) "The Transcendental Dimension of Sartre's Philosophy", in *Reading Sartre*. Ed. J. Webber. Abingdon; New York: Routledge.
Howells, C. (1988) *Sartre: The Necessity of Freedom*. Cambridge: Cambridge University Press.
Kitcher, P. (1990) *Kant's Transcendental Psychology*. Oxford: Oxford University Press.
Jopling, D. (1986) "Kant and Sartre on Self-Knowledge", *Man and World,* 19(1): 79–93.
Leiberman, M. (1997) "The Limits of Comparison: Kant and Sartre on the Fundamental Project", *History of Philosophy Quarterly,* 14(2): 207–17.
Linsenbard, G. (2007) "Sartre's Criticisms of Kant's Moral Philosophy", *Sartre Studies International,* 13(2): 65–85.
Painter, M. (1999) "The Profane Become Sacred: The Protestant Ethics of Kant and Sartre", *Southwest Philosophy Review,* 15(1): 211–17.
Sweeney, K. (1985) "Lying to the Murderer: Sartre's Use of Kant in 'The Wall'", *Mosaic,* 18(2): 1–16.

Part II
Metaphysics

2

Transcendental Unity of Apperception and Non-reflective Consciousness of Self[1]

Sorin Baiasu

1 Introduction

In this chapter, I would like to defend the claim of a deep similarity between Kant's transcendental unity of apperception and Sartre's non-reflective consciousness of self.[2] The claim is not simply of historical interest, although this by itself I think would be sufficient to justify its importance; I take this claim to have also considerable systematic significance. Thus, my motivation for the development of a detailed discussion of Kant's and Sartre's philosophical views is given by the prospect of formulating the outline of a critical ethics which would combine the attractive elements of Kant's and Sartre's theories. I take the deep similarities between Kant's transcendental unity of apperception and Sartre's non-reflective consciousness of self to be important, since they provide the necessary and sufficient condition for an important aspect of accountability.

[1] Part of this chapter was written while an Honorary Guest Research Professor at the University of Vienna, as part of the ERC Advanced Research Project "Distortions of Normativity". I am grateful to the project's PI for making this possible. An early version of this chapter was presented to the 'Jean-Jacques Rousseau' Annual Conference of the Keele Forum for Philosophical Research, which took place in November 2012. I am grateful to Jonathan Weber, who acted as commentator for my paper, and to members of the audience, in particular James Tartaglia, Jochen Bojanowski, Alberto Vanzo and Leslie Sevenson, for stimulating questions and discussion.

[2] I defended this claim quite in detail in my monograph on Kant and Sartre (2011a). Aspects of this claim are also discussed in this volume by Stevenson and Flynn.

One potential problem with such a ground of accountability is that it seems unduly demanding. If the critical ethics that I would like to defend relies for its view of accountability on Kant's transcendental unity of apperception or on Sartre's non-reflective consciousness of self, do I not need very strong assumptions, which are difficult to present and argue for? My answer is negative: the transcendental unity of apperception and the non-reflective consciousness of self are necessary conditions which make possible very basic elements of our cognition and, hence, can be accepted as conditions met by most persons. Yet, assuming that these conditions are indeed very basic and plausible to presuppose, are they really to be found as such in both Kant and Sartre? Is the similarity between them with regard to these notions not breaking when we investigate the nature of these conditions? My answer is again negative: for both Kant and Sartre, necessary conditions for the possibility of the cognition of phenomena are also necessary conditions for the possibility of the existence of phenomena, given that these conditions are constitutive of phenomena.

Both these answers have recently been challenged (Onof 2013). In this chapter, I aim to consider and reply to two objections. These are objections which seem to point to further general and significant issues, particularly in debates on Kant, namely, the issue of non-conceptual content and that of the nature of transcendental idealism. Hence, in addition to responding to these two objections, this chapter also aims to draw some implications for the issues of non-conceptual content and transcendental idealism. Before formulating and addressing these objections (in §§ 4–6), however, I would like to start (in the next section) with a discussion of comparative methodology, a discussion which I take to be essential in a chapter focusing on the comparison of Kant's and Sartre's works.

2 Methodology

A comparative discussion of Kant's and Sartre's philosophies raises an immediate methodological worry. Whether, with regard to a specific topic, one tries to show that Kant and Sartre are similar or different, one will compare particular interpretations of their works. Hence, preferring one interpretation of, say, Kant, to another one is going to lead to distinct conclusions concerning the similarities or differences between the authors compared. Unless there is some reason to adopt some interpretations, rather than some others, it will not be possible to defend convincingly the conclusions of the comparison; suppose Kant and Sartre are similar in some respect, when Kant is understood according to interpretation IK_1

and Sartre, according to IS_1; if, on the basis of some other interpretations (IK_2 or IS_2), they turn out to be different in the same respect, then, given that Kant and Sartre may plausibly be interpreted in various ways, the comparison can hardly be said to illuminate anything.[3]

Let us call this the 'Correct Interpretation' problem and assume there is a way to avoid it (as I will mention below, I think there is); an additional methodological worry emerges then: different authors may use the same philosophical terms with different senses, senses which usually depend on the contexts of their philosophical thoughts. Comparing these authors' claims when they make use of such concepts may be a hopeless enterprise. Thus, it may well be that, although the claims are similar, given that they are formulated in terms which are fundamentally different, the apparent similarity hides in fact very deep differences. Think, for instance, of some potentially similar claims of Kant and Sartre, where words like 'phenomena' or 'freedom' would be used. The plausible implication would be that, given the distinct ways in which these terms are understood by the authors compared, the 'similar' claims may at best be very different and at worst they should not be compared, since they talk about very different things. Let us call this second issue the 'Appropriate Debate' problem.

These two methodological worries are I think a stumbling block that may in part account also for the difficulty of undertaking comparative studies more generally. Specifically in relation to Kant and Sartre, I have discussed these worries in detail elsewhere (2003: 22–4). For the purpose of this chapter, I will mention briefly my methodological assumptions, which are meant to offer at least in part an answer to the Correct Interpretation and Appropriate Debate problems. Concerning the first, although I think it is possible to distinguish between better and worse interpretations,[4] for the sake of this chapter I am going to start from a slightly different requirement. As a background for comparison, I am

---

[3] I first became aware of the significance of this problem in a discussion with Alan Montefiore on Kant and Sartre. I am grateful to him for raising it.

[4] I do not want to deny that we can talk about 'better' and 'worse' interpretations in various senses; we may find an interpretation more accurate, or philosophically more interesting, or logically more compelling, than another one. I will stick here to the claim that it is possible to ascertain whether an interpretation is more accurate than another one and hence better in this sense. I am sympathetic to the accounts offered by the Cambridge School, in particular by John G. A. Pocock (1972, 1981, 2009). Providing a defence of this account would require at least another paper. But the strategy I adopt in this chapter will mitigate the effect of not having such a defence already available.

going to assume what I take to be a 'standard' interpretation of Kant, namely, an interpretation which includes most of the claims commentators usually agree to accept as Kant's. In other words, I will start from an as uncontroversial interpretation of Kant as possible. Quite aside from the methodological issue I raised above (the Correct Interpretation problem), we will see that this assumption I make has particular significance within the context of this chapter.[5]

Consider now the second methodological worry – the Appropriate Debate problem. According to this, the starting point of a comparative analysis of two authors, such as Kant and Sartre, may seem most naturally to be given by claims about a particular topic; say, both Kant and Sartre make claims about freedom and these claims may seem to reflect similarities or differences between their philosophies. The general worry, as we have seen, is that the use made by two authors of the same concepts may be misleading: given that the meaning of such key concepts will depend on the contexts of their philosophical works, it is likely that what seems to be a dialogue between these two philosophers on a topic, like that of freedom, is in fact a set of two parallel monologues, each conducted within the framework of its respective author's thought.

My first suggestion here is that the best place to begin a comparison in this case is given, if available, by comments one author makes on their predecessor. In this way, it becomes clear that there is a topic on which it is likely there is a genuine exchange between the two philosophers. If one of the two authors comments (approvingly or not) on what the second has said on a particular topic, then, at least according to the first, there is a shared set of concepts and views that makes it possible for them to agree or disagree.

Second, I think it is important for this comparison to be allowed to run freely. In other words, although there may be a temptation to try to conclude a comparative analysis by pointing to differences between the two authors' thoughts (especially, as in the case of Kant and Sartre, when these authors are separated by almost two hundred years), this is better to be resisted until a genuine point of disagreement is reached. To put it differently still, although it is very plausible that some, and perhaps very

[5] This still leaves open the question concerning the interpretation of Sartre on the basis of which I will argue here. As elsewhere (2001, 2003, 2010 and 2011a), I am going to offer an interpretation that I claim is accurate. In fact, since this chapter will be concerned with two objections to my interpretation of the relation between Sartre's view of (self-)consciousness and Kant's account of transcendental apperception, I am indirectly defending as accurate both my interpretations of Kant and of Sartre (although, to simplify my task, for Kant I only claim the interpretation is generally palatable in the context of existing literature).

significant, differences exist between these authors' accounts, if these differences do not emerge from the comparative analysis undertaken on the particular topic under discussion, there is no actual requirement to strive to find such differences, especially for a comparison between Kant and Sartre. This is for the following reason.

In the case of Kant and Sartre, the starting point of the comparison will be given by comments Sartre makes on Kant's philosophy and, more exactly, since Sartre is generally very critical of Kant, by Sartre's objections to Kant. Yet, if the starting points are objections formulated by one philosopher to the other, then, given that the objection presupposes a particular interpretation of the ideas to be objected to, two main possibilities emerge. Thus, first, the interpretation of the position to be objected to may be accurate; this will indicate a difference of views between Kant and Sartre, whether or not Sartre's objection turns out to be strong. In this case, therefore, the comparative analysis yields 'naturally' an account of some of the differences between the two philosophers.

If the objection is addressed to a position, which is a misconstrual of Kant's views, then it may turn out that Sartre objects to a view that Kant himself rejects, in which case we identify in this way a similarity between their views: they both object to the same view. Nevertheless, given that what is similar is that they both reject the same view, this similarity may well be underpinned by different perspectives. For instance, both Kant and Sartre may reject realism, even if one does so from a transcendental idealist perspective, whereas the second, from the perspective of an existentialist ontology. In this case, concluding with an account of some of the differences between Kant and Sartre is no longer a requirement, because the similarity identified allows for differences.

In what follows, I will focus on what I take to be a similarity between Kant's transcendental unity of apperception and Sartre's (self-)consciousness or consciousness (of) self. I will briefly outline the argument in support of this conclusion, and then I will formulate two objections that were recently raised to my argument (Onof 2013). In the remainder of the chapter, I will deal with these two objections.

3 Kant and Sartre on the 'I think'[6]

Following the methodological strategy I presented in the previous section, I begin the comparison between Kant and Sartre with an important objection raised by Sartre. In *The Transcendence of the Ego*, Sartre

[6] This section relies on chapters 1 and 2 of my *Kant and Sartre* (2011a).

starts by quoting approvingly Kant's famous claim concerning the possibility of the 'I think'; thus, according to Kant:

> The *I think* must be *capable* of accompanying all my presentations. For otherwise something would be presented to me that could not be thought at all – which is equivalent to saying that the presentation either would be impossible, or at least would be nothing to me (B131–2).[7]

Sartre notes accurately that, on Kant's account, the 'I think' must be *capable* of accompanying my presentations. In other words, it is important to note that Kant does not talk about an 'I think' that would *actually* accompany all my presentations; for some of my presentations, there might be no 'I think', so it is only the possibility of the 'I think' that is regarded as necessary. For instance, there might be situations where, absorbed by what I see (say, a beautiful landscape), I forget about me, as it were; I see the landscape whose beauty absorbs me and what I see is the landscape, rather than myself as seeing the landscape. Hence, the presentation of the landscape that I form is not accompanied by an 'I think'.

Yet, by reflecting on what I am doing, I can bring back the 'I think' and I can then see myself as seeing that landscape. Hence, as Sartre says, in the quotation above Kant focuses on the possibility of the 'I think', not on its *de facto* existence (*TE*: 13–14). Moreover, as Sartre acknowledges, Kant is interested in the *necessity* of the possibility of 'I think' – he regards this possibility as a necessary condition of experience. I will come back to this important Kantian claim.

Nevertheless, Sartre makes it clear that he himself is interested in another type of question; instead of focusing, as Kant does, on *de jure* questions (what *must* be the conditions which make possible this phenomenon or other), he is concerned with *de facto* problems, such as: *Does* a presentation which is not accompanied by the 'I think' (for instance, when I am absorbed by a landscape's beauty) undergo any change when it becomes so accompanied? *Are* presentations unified by

[7] The following abbreviations are used for works of Kant and Sartre, respectively: *Prol*, for *Prolegomena zu jeden künftigen Metaphysik* (AA 04) and AA, for Kant's *Gesammelte Schriften* (1900ff); *TE*, for *La transcendence de L'Ego*. In references, abbreviations will be followed by the volume and page number from Kant's AA. References to the *Critique of Pure Reason* will follow the A (first edition), B (second edition) convention. For Sartre, abbreviations will be followed by page number of the French edition used. Translations used are listed under References.

the 'I think' or is there an already existing unity of presentations that unifies them (*TE*: 15–16)?

The critical character of Sartre's remarks on Kant becomes visible only shortly after this and in relation to the second *de facto* question he raises. Sartre objects to the formal character that the *I* is supposed to have in Kant's account. To see the link between the formal character of the *I think* and the unity of presentations that seems to be connected in some way with the *I think*, consider Kant's account of the thinking I (or what he also calls the "soul") in the Third Paralogism, the Paralogism of Personality.

According to Kant, "What is conscious of the numerical identity of itself in different times is to that extent a *person*" (A362). Consciousness of the numerical identity of itself in different times is a necessary condition of cognition, since without the consciousness that my impression at *t* and my impression at *t+1* are the impressions of the same, numerically identical person, I cannot synthesize impressions; for instance, without that identity, I cannot claim that, say, I had first the impression of red, which then turned out to be that of a particular flower. If I am not conscious that the sensation of red (at *t*) and the perception of a flower (at *t+1*) are presentations of the same person (*my* presentations), then I cannot synthesize them in my perception of a red flower.

On Kant's account, however, this consciousness of the numerical identity of myself is not consciousness of a particular feature of mine. I do not say, for instance, that evidence of my identity, as the person who has a sensation of red and who perceives a flower, is given by the fact that the sensation and perception were both had by an optimistic person. For no matter how fundamentally this feature (say, optimism) were inscribed in my character, it would in principle be possible to think of myself as a pessimist and, hence, to think of myself as identical in a more general sense than that given even by general features of character, such as optimism and pessimism.[8] I could think of myself as possibly different from the way I am and, hence, in seeing myself as identical, I must see myself as identical in a more general sense than that given by the features which define me as I am now (again, say, as an optimist).

It follows that, this consciousness of numerical identity, which makes possible the synthesis of my presentations and, hence, my cognition, is

[8] And, at any rate, saying that the sensation and perception were had by persons with the same feature (optimism or a more specific feature, enthusiastic optimism) still does not amount to the statement of numerical identity required for synthesis.

a very formal element. Because it is meant to unify all my presentations, including presentations of myself as being in a particular way or other, it is a formal element. Because it makes possible the unification of presentations, which is necessary for cognition, this formal consciousness of identity is an epistemological condition. Since what is identical is a formal sense of myself, we can use the 'I' to refer to this identical formal element. Yet, according to Sartre, "an I is never purely formal" (*TE*: 37).

Hence, on Sartre's account, Kant might think that he identified a formal condition which makes possible knowledge; in fact, however, since an *I* is never purely formal (and, hence, must have as content some feature or other), it cannot be a condition which makes possible an understanding of oneself as independent of that feature. It cannot help an optimistic person imagine herself as pessimistic.

Moreover, for Sartre, given that an *I* would have to have some content, it cannot simply be a formal condition for the possibility of cognition and, hence, it would have to be the result of reflection, the result of being conscious of oneself as having some feature of other; as he puts it, "the I ever appears on the occasion of a reflective act" (*TE*: 36). Because not all experience is reflective (that is, not all experience is accompanied by the 'I think'), but all experience is conscious, the *I*, for Sartre, need not be in consciousness.

When the *I* emerges with the 'I think' through reflection, it is part of reflective consciousness; non-reflective consciousness is consciousness of something in the world, something distinct from non-reflective consciousness itself, and something that does not have an *I*. By contrast, reflective consciousness is consciousness of an *I*, and, as an object of consciousness distinct from consciousness itself, it must be considered as part of the world. Given that, whether reflective or non-reflective, consciousness has no *I*, it must be considered as impersonal. Sartre calls it an impersonal transcendental field, and this, Sartre adds, is in a sense a "*nothing*" (*TE*: 74).

This objection to Kant that Sartre formulates in *The Transcendence of the Ego* is the starting point of my comparative analysis of Kant and Sartre. But I have said that I would return to the Kantian claim that the possibility of the 'I think' is a necessary condition of experience, a fundamental epistemological condition of possibility. In his discussion of the Third Paralogism, although Kant criticizes the rationalist claim that we can have cognition of a person's identity starting from the consciousness of numerical identity in different times, he thinks that we can save a weak sense of personal identity. This sense of identity, he says, "is, indeed, needed and sufficient for practical use" (A365). If this sense of identity, which is necessary and sufficient for a certain notion

of responsibility and accountability, is given by the consciousness of numerical identity in different times, then one question is whether it is not too demanding.

If it turns out that, in order to make sense of moral responsibility and accountability, we need to assume the necessity of a very demanding notion of identity, then Kant's claim becomes implausible. After all, it seems evident that moral accountability applies equally well to all persons,[9] irrespective of the degree of sophistication of their views of identity. The notion of identity required as a necessary condition of responsibility is not very demanding, if it can be viewed as met even in the case of an individual who is able to have a very simple perception, such as a sensation of red.[10]

These comments provide the necessary background for two interesting objections to the comparative analysis I defend. The next section will focus on these objections.

4 Two objections

In a generous discussion of my arguments (Baiasu 2011a), Christian Onof focuses in particular on the comparison of what I have called in a general way, in the previous section, Kant's and Sartre's views of the 'I think' (Onof 2013). He correctly notes an implication of the methodological requirement that I spelled out as the second methodological worry in Section 2 of this chapter – the issue of an Appropriate Debate. This methodological worry had to do with the use by Kant and Sartre of similar concepts. By using similar concepts to make similar claims, Kant and Sartre seemed to defend similar views; yet, a closer look at these concepts may indicate that they are used with distinct senses by each author and, therefore, the authors' apparently similar claims are in fact very different.

---

[9] I do not mean all human beings, but something like all moral agents, and we do consider many individuals under this category.

[10] Although this weak notion of identity is necessary for moral accountability, it is not yet sufficient. It would be implausible to suggest that the capacity for having a sensation is a sufficient indication of moral accountability. As I show in more detail elsewhere (Baiasu 2011a: §11), this weak identity is only sufficient for the person's *acceptance* of the status of moral agent, but not for the correct *attribution* of this status to her. It is in this weak sense of acceptance that I claimed, earlier in the chapter, that transcendental unity of apperception and pre-reflective consciousness of self stand for a necessary and sufficient condition of accountability.

Yet, Onof adds, we may also have situations where concepts which are different in Kant and Sartre turn out to be similar. For instance, my claim that Kant's transcendental apperception plays a similar role to that played in Sartre by (self-)consciousness (or non-positional consciousness of self) is of this type. Indeed, the focus in the previous section was precisely on the background necessary for my attempt to defend this similarity.

Onof correctly identifies as one of my concerns that of establishing a notion of a person's identity over time that is sufficient for her status as moral agent.[11] This is both to account for Kant's claim that the formal identity of the transcendental unity of apperception is sufficient for practical purposes and to pursue the initial aim with which I undertook the comparison of Kant and Sartre, namely the aim of rediscovering a neglected type of critical approach in ethics, a type of ethics, however, which would also need to account for the agents' moral status.

Onof notes that, on my reading of Kant's B131–2, I acknowledge Kant's claim that the 'I think' need not accompany any given presentation. Moreover, he correctly notes that I then draw a parallel between this 'I think', which I interpret as an epistemological condition to which Kant's transcendental unity of apperception answers, and Sartre's 'I think'. The Kantian possibility of the 'I think', which is a pre-reflective form of self-consciousness, plays, I then claim, a role similar to that played by Sartre's (self-)consciousness. Moreover, as I have mentioned in the previous section, I attempt to show that both the transcendental unity of apperception and non-reflective consciousness of self, as conditions of knowledge, are not very demanding. If having even a simple sensation presupposes that the epistemological condition is met, then the condition is not very demanding.

According to Onof, however, this leads me to "make some controversial claims about the conditions for having a sensation, namely that this requires a synthesis of presentations" (2013: 323). Thus, he continues, if I take this synthesis to be an epistemological condition, then I am appealing to "a Strawsonian identification of sensation with sense-data, and to the correct claim that the latter requires the ability to differentiate" between sense-data and, hence, requires self-consciousness (2013: 324). Yet, if sense-data require self-consciousness, this is because they are misleadingly conceived of as objects, in Kant's sense of the notion (A92/B125); by contrast, according to Onof, in fact, for Kant, "sensation is to

[11] However, as I have mentioned in the previous footnote, my concern is to establish only such a sufficient condition insofar as we focus on the agent's own acceptance of the status of moral person.

be understood as a pre-objective ingredient in the construction of the object" (2013: 324).

The second objection which I will consider in this chapter is that, from the way I discuss these issues, it is unclear whether what I am actually interested in are not ontological conditions, "e.g. what is required to have a sensation"; moreover, Onof suggests, I seem "to conflate [this ontological condition], again in Strawsonian fashion, with Kant's notion of *Erfahrung*" (2013: 324). Evidence of my interest in ontological conditions, Onof suggests, would be my claim that there is a similarity between Sartre's (self-)consciousness and Kant's transcendental unity of apperception, which, for Onof, would represent "an attempt to give a Kantian spin to such an ontological condition" (2013: 324). But, if this is so (that is, if I am interested in ontological conditions), on Onof's account, I should not look at the Transcendental Deduction, which does not deal with ontological conditions.[12]

Let me try to spell out a bit further these two objections, in order to present more explicitly the challenge they raise for my argument.[13] The presupposition of the first objection seems to be that any "pre-objective ingredient in the construction of the object" cannot itself require a synthesis of presentations. The implication is that only 'constructed' objects require a synthesis, since this is how their "construction" (Onof 2013: 324) is possible. A synthesis of presentations, which also makes possible discrimination between these presentations, needs conceptual input from the understanding. But it is not completely clear that Kant regards sensations as needing or even as able to accommodate conceptual input – and, Onof adds, this by itself is an issue of on-going dispute for those involved in the debate concerning non-conceptual content. Hence, on Onof's account, to assume that sensations do have such a conceptual input, I would need to regard them as already 'constructed' objects and, hence, as "Strawsonian sense-data".

---

[12] Finally, Onof notes, all this makes "very puzzling" my attempt to make sense of the Kantian distinction between subjective and objective unities of consciousness; this I would do in a "questionable" fashion, by an identification of the subjective unity with the empirical unity of apperception, and, yet, subjective unity concerns how the manifold is "given for … combination" (B139) and cannot therefore require a synthesis. (2013: 324). I do not think this is the case, but Onof does not provide further argument here; he only refers to his 2010 paper. Given the limited scope of this chapter, a discussion of this third objection will have to be postponed for another occasion.

[13] Great help to clarify the nature of these objections was provided by further email correspondence with Onof.

One assumption of the second objection is that ontological conditions are conditions for the existence of things in themselves.[14] By contrast, a priori conditions for the possibility of phenomenal objects, in particular concepts of the understanding, are not ontological, because they *determine* the object of experience, whereas the existence of the object cannot determine the object in addition to how it is determined by its epistemological conditions. If the object's existence would be determined by the a priori conditions for the possibility of the object, then we would end up with subjective or Berkeleyian idealism. Given that I talk about the a priori conditions for the possibility of objects as ontological or metaphysical conditions and that I assume that sensations, too, are constituted by such conditions, I must conflate sensations with experience in a (as already indicated in the first objection) Strawsonian fashion, and I must attempt to draw the comparison between Kant's transcendental unity of apperception and Sartre's non-reflective consciousness of self in order to give a "Kantian spin" to such an ontological condition (Onof 2013: 324). The implication is that Sartre's (self-)consciousness would represent an ontological condition in the sense specified.

But it is not clear these objections have much force. In the next two sections, I will argue that both of them rely on questionable assumptions.

5 Response to first objection

Consider first the suggestion that a sensation, as an ingredient of the Kantian object, is an element of the synthesis necessary for the constitution of the object, rather than being the result of such a synthesis. As I have mentioned in the previous section, the assumption on which this objection relies is that a pre-objective ingredient in the construction of an object cannot be the result of a process of synthesis. But it is unclear why this must be so. More exactly, it can be granted that some pre-objective ingredients may not be the result of a synthesis, but it is unclear why we should accept that this would be the case for all of them.

The example I make reference to is that of a sensation of red. I claim that such a sensation involves a synthesis and, hence, presupposes at work the transcendental unity of apperception or the non-reflective consciousness of self. Thus, in order to have a sensation of red, one needs to be able to discriminate at least between what is red and what is not red. Hence, in order to synthesize these presentations, the transcendental

[14] Onof (in email correspondence) points to this use of the expression by Henry Allison in the first edition of his *Transcendental Idealism* (1983).

unity of apperception is a necessary condition. Discriminating is an activity which requires synthesis and the activity of the understanding. The crucial question is, therefore, whether we can accept, on the basis of Kant's account, that sensations have conceptual content. It is for this reason that Onof thinks the debate is linked to the important current debate concerning non-conceptual content.

Consider the following short discussion of Kant's notion of sensation [*Empfindung*]:

> A sensation arises out of the faculty of representation being affected by the presence of an object (CPR A19/B34). It is described as the 'matter' of appearance and distinguished from perception which is sensation accompanied by consciousness, although it too is occasionally described as the 'matter' of perception. It is also described (in ID §4) as the matter of sensibility, which is complemented by its form or 'co-ordination'. (Caygill 1995)[15]

Sensation is presented here first by reference to Kant's definition at A19/B34: "The effect of an object on our capacity for presentation, insofar as we are affected by the object, is *sensation*." Hence, sensation is generated by our capacity of representation as an effect of its being affected by an object. This effect is considered the matter of appearance. A distinction is then introduced between sensation and perception, the latter being sensation accompanied by consciousness. This suggests sensation would not be accompanied by consciousness. Finally, the short description of sensation distinguishes between the form and matter of sensibility, a distinction present also in the *Critique of Pure Reason*:

> Whatever in an appearance corresponds to sensation I call its *matter*; but whatever in an appearance brings about the fact that the manifold of the appearance can be ordered in certain relations I call the *form* of appearance. Now, that in which alone sensations can be ordered and put into a certain form cannot itself be sensation again. (A20/B34)

This suggests that a synthesis will be present at the level of the appearance, but not at the level of sensations; sensations are the matter of appearance and their manifold is synthesized by the form of appearance. Given that sensations are not supposed to be accompanied by

[15] "CRP" refers here to Kant's first *Critique*, whereas "ID" to Kant's inaugural dissertation, *De mundi sensibilis atque intelligibilis forma et principiis* (AA 02).

consciousness, it becomes difficult to see them as the result of a synthesis involving the transcendental unity of apperception or the non-reflection consciousness of self. This seems to support very strongly Onof's first objection to my account. To be sure, it is not so much the comparison I draw between Kant and Sartre that is undermined; as I mentioned in the previous section, my aim in using the example of the sensation of red was to show that the necessary condition of a transcendental unity of apperception is not very demanding. It might be that the same argument can be constructed starting from an appearance of red, as opposed to a sensation of red.

Irrespective of this, the objection would still point to a problem in my account of Kant; thus, first, as I have said in Section 2, in relation to the Correct Interpretation problem, my claim was that the account of Kant I use in the comparison is largely uncontroversial. By contrast, if correct, the objection would show that my interpretation would be controversial, to say the least, and in particular from the perspective of the following two implications.

First, Kant's discussion at A20/B34 takes place at the level of sensibility; hence, even if I were to focus with my example on an appearance, it would seem that the transcendental unity of apperception, although present, could not on this basis alone be considered as a condition of cognition. This is because, for Kant, cognition requires both sensibility and understanding, and an appearance – as confined to sensibility and lacking the input of the understanding – does not amount yet to cognition. It is in this sense also that we can say that sensations and appearances are not objective presentations.

Second, however, it is doubtful we can even talk about an appearance of red, or about having a sensation of red: an appearance of red would indeed require a distinction between appearances of red and appearances of different colours, and this distinction would be more than the a priori forms of sensibility could provide. As I have mentioned, this kind of discrimination is the result of the activity of the understanding. Moreover, in order for me to talk about having a sensation, I would need to experience this sensation.[16] Since experience requires both sensibility

[16] I cannot say I am having a sensation, if I am relying on indirect evidence; say, a colour-blind person, who cannot distinguish between red and blue, but whose blood pressure increases when she is surrounded by red objects, would be able to say that she is surrounded by red objects and point to the evidence of the higher blood pressure, but she would not say that she has a sensation of red. More could be said here, of course.

and understanding, and since sensations are merely a basic element of sensibility, I could not have a sensation of red.

Be that as it may, however, Kant does talk about having a sensation of red. Consider the following passage, in the *Prolegomena*, where Kant responds to one version of the objection that his philosophy is nothing but a version of traditional (subjective) idealism:

> I would very much like to know how then my claims must be framed so as not to contain any idealism. Without doubt I would have to say: that the representation of space not only is perfectly in accordance with the relation that our sensibility has to objects, for I have said that, but that it is even fully similar to the object; an assertion to which I can attach no sense, any more than to the assertion that the sensation of red [*die Empfindung des Rothen*] is similar to the property of cinnabar that excites this sensation in me [der diese Empfindung in mir erregt]. (*Prol* 4:290)

Kant responds here to the worry that, in considering space and time as a priori structures of sensibility, rather than as properties of things as they are in themselves, he would reduce objects to representations in our minds. In response, he notes the distinction between sensibility and the objects of our sensibility; moreover, he notes that space, as one of the a priori forms of this sensibility, is in perfect accordance with the link between sensibility and the objects that affect sensibility.

In other words, the sensations that are produced when objects affect our sensibility are the content that is organized by the forms of sensibility (space and time). Space is therefore in perfect accordance with the link established by sensation between sensibility and the objects affecting it. Yet, since this claim is criticized as idealistic, Kant suggests a more radical one: that space is fully similar to the object. It is this more radical claim that he deems absurd and compares with the claim that the sensation of red is similar to the property of cinnabar that excites this sensation in me. Since colour is not a property of cinnabar, it is strange to say that the sensation of colour would be similar to a property in the object that produces the sensation. The sensation of colour is perfectly in accordance with the relation that our sight has to the object under appropriate conditions, but it is not fully similar with the property of the object that affects me and produces the sensation.

I have discussed the issue of the distinction between Kant's transcendental idealism and traditional idealism at length elsewhere, and

I will return to this in my response to the second objection.[17] Here, the discussion of idealism provides the context in which Kant talks about a sensation of red, and my focus is on this. It should first be noted that, contrary to the previous conclusion, it seems that it is perfectly legitimate to talk about having a sensation of red within the Kantian framework. This, however, does presuppose that the person who has this sensation can differentiate between a sensation of red and a sensation of another colour. This, in addition, suggests that a synthesis of presentations performed with the help of the understanding is presupposed by the sensation of red and, hence, that a sensation of red is not simply a pre-objective element from which phenomena are constituted.

Kant takes this expression to refer to something produced by a property in the object that affects sensibility. Hence, he uses the expression in the usual sense: as an element of the experience of cinnabar. Yet, as I have mentioned above, a sensation of red presupposes both sensibility and understanding, and can be seen as an element of cognition too (even if only corresponding to an inner experience).

This is puzzling, however, because, as I have mentioned, one of the implications drawn earlier in this section was that I cannot have a sensation of red, precisely because sensation, on Kant's account, is an element of sensibility (and even one independent from consciousness), and I can only have an experience of red, which involves both sensibility and the understanding. But this puzzle suggests also its own solution: when Kant talks about a sensation of red, he does mean an *experience* of a particular colour, one element of the experience of an object (cinnabar). By contrast, when he talks simply about sensations, he refers to the effect produced by the way sensibility is affected by an object, that is, he refers to the sensible element of an object's experience.[18]

I conclude therefore that Onof is right that a sensation of red is an element of the experience of an object, but this element is not pre-objective. There seems to be at work here an ambiguity over the concept of an element, which can be understood as a constitutive element of experience (sensation and concept, for instance) or as a part of an experience (a sensation of a particular colour or shape). To talk about a sensation of

[17] See my texts (2013a and 2013b).

[18] What Kant denies is that colour is a property of the object, just as he denies that space is a property of things in themselves; but, he claims, this is not to deny the existence of the object. Colour attaches to the sense of vision as a modification, but this does not deny the existence of a property of the object that affects the sense of vision and produces the modification that corresponds to the sensation of colour (*Prol* 4: 289).

red is to talk about an element of experience in the latter sense, a sense which, *pace* Onof and Schulting,[19] makes no Strawsonian misidentification of sensation and sense-data; or, rather, since Kant himself talks about a sensation of red produced by cinnabar and must mean an aspect of the experience of red, the identification of determined sensations with sense-data is as Kantian as it is Strawsonian.

6 Response to the second objection

According to the second objection, when I talk about the transcendental unity of apperception and non-reflective consciousness of self as epistemological conditions of cognition, it is unclear whether, in fact, what I have in mind are not ontological conditions. But, if I refer, again (according to Onof) in a Strawsonian fashion, to ontological conditions, I should not discuss Kant, whose conditions are epistemological; instead, I should perhaps discuss Sartre. Comparing Kant's transcendental unity of apperception and Sartre's non-reflective consciousness of self would be precisely an attempt to give a Kantian spin to such ontological conditions.

I think one assumption here is that, by 'ontological conditions', I would refer to conditions which make the existence of things in themselves possible. One can of course follow here Allison and understand an ontological condition in this sense.[20] Yet, because both Kant and Sartre deny things in themselves (the former epistemically, whereas the latter, metaphysically), talking about ontological conditions in this sense would not be relevant for either Kant or Sartre (although it would be relevant for their refutation of various positions, which rely on their epistemic or ontological possibility). By contrast, the notion that I use refers to the conditions which are necessary for the existence of phenomena. Alternatively, I can say that these are ontological conditions of empirical reality.

---

[19] In formulating his objection as pointing to a Strawsonian misidentification, Onof acknowledges Dennis Schulting (Onof 2013: 328 n. 1).

[20] According to Allison, an epistemic condition is "one that is necessary for the representation of an object or an objective state of affairs" (1983: 10). An ontological condition is a condition "of the possibility of the being of things", and, because "the being of things is here contrasted with their being known, an ontological condition is, by definition, a condition of the possibility of things as they are in themselves (in the transcendental sense)" (1983: 12). A similar distinction is used also in the second edition of Allison's book (Allison 2004).

However, I think the objection merits further discussion, because there seems to be an implication in the way Allison defines ontological conditions, an implication according to which they can *only* be defined as conditions necessary for the existence of things in themselves. Thus, according to him:

> It is […] important to distinguish epistemic from ontological conditions. Since the being of things is here contrasted with their being known, an ontological condition is, *by definition*, a condition of the possibility of things as they are in themselves. (Allison 1983: 11 – my emphasis, SB)

The suggestion here is that, if I talk, as I actually do, about ontological conditions, I must have in view conditions of the possibility of things as they are in themselves. If this is correct, then Onof would be right to suggest that, in talking about ontological conditions, I must refer to the existence of things as they are in themselves. To be sure, I do distinguish between ontological and epistemological conditions, but I also claim that, for Kant, at least some of them refer to the same things (in particular, the categories or the a priori concepts of the understanding). It is probably also this additional claim, which produces confusion and suggests that, although I claim that the transcendental unity of apperception (in Kant) and the non-reflective consciousness of self (in Sartre) are epistemological conditions, I might in fact have in view ontological conditions.

But what might be the reason for Allison's suggestion that, by definition, ontological conditions must refer to conditions of things in themselves? The argument I will consider is the following: even if we were to take ontological conditions to refer to the necessary conditions for the existence of phenomena, as I actually claim to do, one would ultimately have to refer to the conditions of things in themselves, because, without considering things in themselves, phenomena would be ontologically reducible to structures of our mind and, hence, to *ideal* entities. If we are to account for the *reality* of phenomena, we need ultimately to presuppose that phenomena are grounded in, or supported by, things in themselves and, hence, the ontological conditions of phenomena would ultimately be the conditions of things in themselves.

As I have argued elsewhere (2011b and 2013b), however, this is the kind of argument that motivates the standard criticism of Kant as advancing nothing more than a traditional form of idealism, an argument to

which, in the previous section, we have seen that Kant responded in the *Prolegomena*. The argument seems to rely on the assumption that phenomena are either 'really real' (when they are supported by things in themselves) or subjective (if it turns out things in themselves do not exist). Given that we cannot establish whether things in themselves exist or not, and given that simply to assume they exist is a strong assumption, it seems safer to conclude phenomena are subjective and Kant's transcendental idealism, simply a form of traditional idealism.

Similarly, the assumption that phenomena are either 'really real' or subjective suggests that they can only have ontological status when they are grounded in things in themselves and, hence, their ontological conditions would ultimately be conditions of things in themselves. If this were so, then Onof's second objection would apply, even if what I actually had in mind were conditions of the existence of phenomena. And, yet, I think the argument in support of Allison's view of ontological conditions does not hold.[21] Thus, for Kant, the distinction between the subjective and objective character of what we might call 'epistemic entities' (such as, claims or assertions, but also concepts or principles) can be drawn in two ways – empirically or transcendentally.

Thus, empirically, inner experiences in consciousness or 'ideas' (in the way in which traditional idealism understands them) are subjective – they are 'in the mind' and, hence, they are not real. By contrast, a priori intuitions (space and time), as well as the phenomena, are empirically real – they are not simply in the mind (or inner) and, hence, are empirically objective. For instance, an object in space is outside me and, hence, empirically outside my mind; it is independent from my mind, as it cannot be created and destroyed mentally in the way in which I can do this with a particular thought through imagination. Transcendentally, however, the a priori intuitions as well as the other a priori structures of the mind are subjective, in the sense that they depend on the mind; from a transcendental perspective, objective are only things in themselves – these are independent from the mind.

In other words, phenomena are neither inner states of mind, nor things in themselves; they are empirical realities, which do have an ontological status, even if not as strong as that of things in themselves.

---

[21] In fairness, I should specify that this in fact is not so clearly Allison's claim as Onof's interpretation, since Allison qualifies the definition of ontological conditions in a crucial way: "an ontological condition is, by definition, a condition of the possibility of things as they are in themselves (*in the transcendental sense*)" (1983: 11 – my emphasis, SB). The relevance of this will become clear shortly.

Ontologically, phenomena can be distinguished from states of mind or ideas. If this is so, then an investigation into the ontological conditions of phenomena need not force us to investigate the ontological conditions of things in themselves.[22] The transcendental unity of apperception and the non-reflective consciousness of self are necessary conditions of cognition and, hence, are epistemic conditions of phenomena. But they are not only necessary conditions which make possible cognition of phenomena, they are also ontological conditions of phenomena, insofar as they are constitutive of phenomena. This is in the same way in which a priori intuitions and concepts are constitutive of phenomena. A priori intuitions and concepts, as well as the transcendental unity of apperception and the non-reflective consciousness of self, are empirically *real*. Although they are ontologically distinct from, and weaker than, things in themselves and although we have no way of deciding whether things in themselves exist, we are still able to distinguish between phenomena (say, external objects) and illusions. Hence, an investigation into the ontological conditions of phenomena need not ultimately lead to an investigation into the ontological conditions of things as they are in themselves.

To be sure, the question of what confers (empirical) reality to these a priori structures of the mind and the extent to which, without being supported by things in themselves, this reality does not collapse to become as weak as inner states of mind – these remain open questions here. Moreover, given that the ontological status of phenomena seems unstable (between the real reality of things in themselves and the illusion of inner thoughts), Kant's transcendental idealism seems permanently under the threat of the return of the exclusive disjunction mentioned above: a priori structures of the mind are either properties of things in themselves or they are as illusory as inner states of mind.

Yes, precisely because we cannot say that things in themselves either exist or do not exist, but we can say that we are able (or should be able) to distinguish between the empirical reality of objects and the unreal

[22] I have said (in the previous footnote) that Allison qualifies the definition of ontological conditions by specifying that they represent conditions of things as they are in themselves in the *transcendental sense*. This shows clearly that Allison has in view here the possibility of considering things as they are in themselves in the empirical sense, and this means to consider them as empirical realities, as phenomena. This qualification suggests that he need not assume that an investigation into the ontological conditions of phenomena is ultimately an investigation into the ontological conditions of things in themselves.

character of our states of mind, we should hold fast to the empirically real character of phenomena and worry about the origin of this reality and the existence or non-existence of things in themselves as secondary (although very important) questions. The second objection raised by Onof can therefore be answered by pointing to the legitimacy of this sense of ontological condition.

7 Conclusion

This chapter examined two objections to my claim that there is a deep similarity between Kant's transcendental unity of apperception and Sartre's pre-reflective consciousness of self. The first objection focused on my argument that the transcendental unity of apperception and the pre-reflective consciousness of self represent conditions of accountability which are not very demanding. According to my argument, they are epistemological conditions necessary for the synthesis presupposed even by simple particular sensations, for instance, a sensation of red. Yet, the objection criticizes my assumption that a sensation would have an objective character that would require a synthesis and, hence, the contribution of the understanding. By contrast, according to this objection, a sensation is only a component of experience, a pre-objective element that does not need, and in fact cannot presuppose, a synthesis.

In response to this objection, I have showed that there is an equivocation over the notion of an element of experience, which can mean both an objective part of experience and a pre-objective component of experience. According to the first objection, my argument takes a pre-objective component of experience (sensation) and claims to be the result of a synthesis made possible by the transcendental unity of apperception or the pre-reflective consciousness of self, a synthesis which is only presupposed by an experience or an objective part thereof. Yet, my argument refers to such an objective part of experience by talking about the synthesis presupposed by a sensation of red a person might have. Moreover, given that my argument refers to such an element of experience, it does not have implications for the recent debate on non-conceptual content.

The second objection questions my claim that the transcendental unity of apperception or the pre-reflective consciousness of self would be epistemological conditions and suggests that I am in fact focused on ontological conditions of sensation and that such conditions are not

discussed by Kant, but are investigated by Sartre. My attempt to indicate a similarity between Kant's transcendental unity of apperception and Sartre's pre-reflective consciousness of self would be an attempt to give a Kantian spin to such an ontological condition.

In response to the second objection, I have questioned the assumption that ontological conditions of phenomena must refer to necessary conditions for the existence of things in themselves. I have argued that such an assumption is at the basis of standard arguments which claim to show that Kant's transcendental idealism is at the end of the day not much different from traditional idealism. I have shown that such a claim ignores the significant emphasis Kant puts on the fact that his account is appropriately presented as both transcendentally ideal *and* empirically real. By restricting the notion of ontological conditions to the necessary conditions which make possible the existence of empirically real phenomena, I have shown that the transcendental unity of apperception and the pre-reflective consciousness of self, as constitutive of empirically real phenomena, are ontological and epistemological conditions of phenomena, and that it is perfectly legitimate to talk about ontological conditions in Kant.

We have seen that, in his attempt to reject the accusation of traditional idealism, Kant offers an analogy. He starts from the standard acknowledgement that colour is not a property of the object itself, but a modification of vision, and he notes that denying this status to colour is not usually taken to imply idealism, since the object affecting our vision is not denied. Similarly, he says, the fact that space belongs to the appearance of an object, rather than to the object in itself, does not imply that there is no object; it only implies that we cannot cognize the object as it is in itself.

It follows that the conditions which make possible our cognition of appearances or phenomena are also conditions which make possible the existence of appearances or phenomena, and these ontological conditions need not be conditions for the existence of things in themselves. Thus, according to Kant, it is perfectly acceptable that

> one could, without detracting from the actual existence of outer things, say of a great many of their predicates: they belong not to these things in themselves, but only to their appearances and have no existence of their own outside our representation. (*Prol* 4: 289)

Asserting the empirical reality of appearances or phenomena is something both Kant and Sartre would be able to do within their philosophical

frameworks. Talking about the transcendental unity of apperception and the pre-reflective consciousness of self as necessary conditions which make possible cognition of phenomena would, I have claimed, be, for both philosophers, compatible with the claim that this transcendental unity of apperception and this pre-reflective consciousness of self are also ontological conditions of phenomena. They are constitutive of phenomena and this need not imply that they would have to refer to things in themselves.

References

Allison, H. E. (2004) *Kant's Transcendental Idealism: An Interpretation and Defense*. Rev. Exp. New Haven, CT: Yale University Press.

—— (1983) *Kant's Transcendental Idealism: An Interpretation and Defense*. New Haven, CT: Yale University Press.

Baiasu, S. (2001) *Persons and Politics in Kant and Sartre*. PhD Thesis. University of Manchester.

—— (2003) "The Anxiety of Influence: Sartre's Search for an Ethics and Kant's Moral Theory", in *Sartre Studies International* 9(1): 21–53.

—— (2010) "Kant's Account of Motivation: A Sartrean Response to Some Hegelian Objections", in *Hegel Bulletin* 31: 86–106.

—— (2011a) *Kant and Sartre: Re-Discovering Critical Ethics*. Basingstoke: Palgrave Macmillan.

—— (2011b) "Space, Time and Mind-Dependence", in *Kantian Review* 16(2): 175–90.

—— (2013a) "Introduction: Kant and the British Idealists", in *Collingwood and British Idealism Studies* 19(1): 1–18.

—— (2013b) "Caird on Kant's Idealism: Traditionalist or Revolutionary?", in *Collingwood and British Idealism Studies* 19(1): 19–45.

Caygill, H. (1995) *A Kant Dictionary*. Oxford: Wiley-Blackwell.

Kant, I. (1900ff) *Gesammelte Schriften*. Vols 1–22 Preussische Akademie der Wissenschaften; vol. 23 Deutsche Akademie der Wissenschaften zu Berlin; vols 24- Akademie der Wisseschaften zu Göttingen.

—— (1996 [1781/1787]) *Critique of Pure Reason*. Tr. Werner S. Pluhar. Indianapolis/Cambridge, IN: Hackett.

—— (2002[1783]) "Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science". Tr. G. Hatfield, in *Theoretical Philosophy after 1781*. Eds H. Allison and P. Heath. Cambridge: Cambridge University Press.

Onof, C. (2010) "Kant's Conception of Self as Subject and Its Embodiment", in *Kant Yearbook* 2: 147–74.

—— (2013) "Review: Sorin Baiasu, *Kant and Sartre: Rediscovering Critical Ethics"*, in *Kantian Review* 18(2): 323–8.

Pocock, J. G. A. (1972) *Politics, Language, and Time: Essays on Political Thought and History*. Chicago: University of Chicago Press.

—— (1981) "Virtues, Rights, and Manners: A Model for Historians of Political Thought", in *Political Theory* 9(3): 353–68.

—— (2009) *Political Thought and History: Essays on Theory and Method*. Cambridge: Cambridge University Press.
Sartre, J.-P. (1985[1936-7]) *La transcendence de L'Ego: Esquisse d'une Description Phénomènologique*. Ed. Sylvie le Bon. Paris: Vrin.
—— (2004[1936-7]) *The Transcendence of the Ego: A Sketch for a Phenomenological Description*. Tr. A. Brown. London: Routledge.

3
Kant and Sartre on Temporality

Daniel Herbert

Notwithstanding the enormity of his intellectual debt to the German philosophical tradition, Sartre's remarks concerning Kant are, in *Being and Nothingness*, more often critical than complimentary.[1] Sartre's antipathy to the Critical philosophy is perhaps especially apparent in his discussion of temporality, where Kant is accused of failing to account either for the 'order' of time or for its 'course'[2] Moreover, what must have seemed especially objectionable to Sartre given his pre-eminent concerns with the recognition of human freedom is that, in his view, Kant's treatment of temporality excludes any possibility of spontaneous agency on our part, and therefore commits us to the denial of our fundamental status as autonomous agents.

As a number of commentators have noted, however, there are strong affinities between Sartre's phenomenological ontology and Kant's transcendental philosophy, both of which propose to examine the phenomena of experience in terms of their relation to, and necessary conditionality upon, certain fundamental features of human subjectivity.[3] Indeed, it is because of their common preoccupation with features of human experience which they regard as universal and necessary

[1] I am especially grateful to Sorin Baiasu for inviting me to participate in the 2012 Keele conference on "Kant and Sartre" at which an early version of this chapter was presented, and to Anna Tomaszewska for her extremely helpful comments in reply to the presentation which I gave at that event.

[2] The Kantian distinction between the 'course' of time and its 'order' corresponds to MacTaggart's better-known distinction between the A-series and the B-series of temporal succession.

[3] Gary Cox (2006), Michelle Darnell (2006) and Sebastian Gardner (2009), for instance, have each argued that Sartre may be considered as a kind of transcendental philosopher.

that Kant and Sartre are similarly motivated to examine the temporal structures that condition our perspective upon the world. As such, the particular objections which Sartre raises against certain features of Kant's account of temporality are liable to obscure more fundamental points of agreement between their respective positions.

Kant's account of time is similar to Sartre's in that neither recognizes temporality as a feature of that which is 'in itself', or unconditioned by any relation to the subject's mode of experience and understanding. Hence, for both philosophers, it is only from a human perspective that we may speak of time. According to Kant and Sartre, time is a defining and ineliminable feature of the human condition, and it is in terms of temporality that we are able to determine our experience as meaningful or significant for us. As such, Kant and Sartre are similarly concerned to clarify the ontological status of time and its significance for human experience. In the following, Kant's and Sartre's treatments of temporality are compared and the extent of their agreement shown, while diagnosing the sources of what disagreements there are between their respective positions.

1 The centrality of the representation of time in Kant's Critical project

On February 21, 1772, Kant wrote to his former student Marcus Herz, claiming to have identified "the key to the whole secret of the metaphysics that had until then remained hidden to itself" (Prol. 117). Kant's letter to Herz sketches an outline of the Critical philosophy in the early stages of its development and addresses a question which would ultimately receive its answer in the characteristically Kantian combination of transcendental idealism and empirical realism, namely: "on what grounds rests the reference of what in us is called representation to the object?" (Prol. 117). It is also in answer to this question that Kant distinguishes between intuitions and concepts, two species of representation, both of which, he claims, can be either a priori or a posteriori.

That the possibility of cognitive access to the object rests necessarily upon the independent contributions of both intuitional and conceptual modes of representation is a commitment fundamental to the Critical philosophy. According to Kant, the use of concepts is necessary for us to form representations of general classes, or rather, general criteria of membership of one or another class. Without concepts, then, we would be unable to represent objects as conforming to general laws, or as sharing properties in common with other objects. While concepts

enable us to have predicative thoughts, however, they are incapable, in Kant's view, of securing reference to objects, since the mere availability of norms of membership of a certain class are insufficient to determine whether or not such a class in fact has any members. Knowledge of what criteria an object must satisfy in order for it to count as a member of the class of unicorns, for instance, is inadequate for knowledge of whether any object happens in fact to meet these conditions. In order for it to be possible for us to recognize, in a possible object of cognition, the general properties which mark it out as a member of a certain class, that object must be 'given' to us by non-conceptual avenues, and this, Kant maintains, is the role of sensible intuitions. Whereas concepts are responsible for introducing general patterns into the manifold of representational content, it falls specifically to sensible intuition to initially provide such a manifold for conceptualization.

Hence, for Kant, concepts relate to objects by mediation of intuitions through which an object is immediately given. However, as Kant argues in the Transcendental Aesthetic, that part of the first *Critique* which is devoted to the a priori form of the specifically intuitional dimension of cognitive experience, the human form of intuition is such that objects can only ever be given to us under the conditions of space and time. Nonetheless, Kant maintains that whereas the representation of space is a condition only of our representation of those objects which come to us 'externally', by means of our contingent physical senses, all of our representations must be temporal insofar as they must occupy a place in the temporally extended sequence of one's train of thought, whether or not the representation in question relates to an object to which we could possibly become related by visual, auditory, or other physical means. Hence Kant distinguishes between space as the pure form of 'outer intuition' and time as the pure form of 'inner intuition', and claims that:

> [t]ime is the a priori formal condition of all appearances in general. Space, as the pure form of all outer intuitions, is limited as an a priori condition merely to outer intuitions. But since, on the contrary, all representations, whether or not they have outer things as their object, nevertheless as determinations of the mind themselves belong to the inner state, while this inner state belongs under the formal condition of inner intuition, and thus of time, so time is an a priori condition of all appearance in general, and indeed the immediate condition of the inner intuition (of our souls), and thereby also the mediate condition of outer appearances. (A34/B50)

As such, for Kant, since time is the a priori form of all possible intuitions, any mode of conceptual thought which relates to an object must do so by mediation of representations which are temporal in character. That this applies as much to a priori concepts as to their a posteriori counterparts is readily apparent from Kant's discussion of the Schematism, in which it is argued that conceptual form must receive a temporal mode of representation if it is to gain any purchase upon the phenomena of sensible intuition. Hence, for Kant, the representation of time enjoys an especially privileged status in accounting for the possibility of cognitive access to the object.

However, as Kant famously argues in his transcendental deduction of the categories, sensible intuition puts us in receipt of a manifold of temporally successive phenomenal contents, the possible representation of which is necessarily conditional upon their reduction to synthetic unity by means of a conceptual form originating in the spontaneity of the thinking subject's rational understanding of its own experience. For there to be any cognitive significance to the subject's receipt of representational content by way of spatio-temporal intuitions, Kant maintains, the subject must be able to recognize each of the items of sensible content so delivered as being somehow relatable to one and the same subject, which is identical with itself. Were this not the case, were it not possible for the subject to recognize the manifold contents of sensible intuition as directed upon a single point of unity with which that subject is identical, then there could not be one and the same subject in receipt of a diversity of representations, and each representation would have to relate to a distinct subject. Under circumstances such as these, the cognitive opportunities available to the variety of subjects in question would have to be extraordinarily limited, no such subject being capable of forming representations of anything more complex than a simple spatio-temporally located feeling of a particular warmth, colour or sound, for instance. Clearly, then, this is not the scenario in which the human subject finds itself, and it must indeed be the case that a variety of intuitions of sensible content are able to be recognized by one and the same subject as somehow relatable to that very same subject as upon a single fixed point of purely formal unity.

It is this very recognitional capacity to which Kant refers as the transcendental unity of apperception. The various items of sensible content must be representable as converging upon a single "unity", and this is the very unity which obtains in "apperception", or the subject's reflexive awareness of its own cognition of various sensible contents. Sensible intuition could not contribute to the possibility of cognitive access to

the object were its content given in such a manner that our conceptual understanding could not recognize anything of itself in the manifold of intuitional representation. The intelligibility of those representations issued by sensible intuition is necessarily conditional upon the possible exercise of conceptual modes of representation by means of which the subject may articulate a judgement concerning the content of its own thoughts. As Kant makes especially apparent in the A-Deduction, even the representation of time would amount to nothing, so far as cognitive experience is concerned, were it not subject to conceptually articulable structures of thought. Hence, for Kant, the experience of empirical phenomena as temporal ultimately has as much to do with conceptual understanding as with sensible intuition. As Heidegger came to appreciate in his admittedly heterodox interpretation of the Critical philosophy, the structure of temporal experience is clearly of paramount concern to Kant's project, and is the focus of his attention in understanding the possibility of metaphysics.[4]

2 Kant and the transcendental ideality of time

However, and notoriously so, Kant also argues in the Transcendental Aesthetic that "[t]ime is not something that would subsist for itself or attach to things as an objective determination, and thus remain if one abstracted from all subjective conditions of the intuition of them" (A32/B49). Rather, he claims, "time is nothing other than the subjective condition under which all intuitions can take place in us". (A33/B49) As such, insofar as all possible objects of sensible intuition are subject to the a priori necessary condition of time, Kant is committed to holding that the objects of possible sensible experience are, without exception, transcendentally conditional upon a certain feature of our human mode of cognition. It is therefore in the Transcendental Aesthetic that Kant first argues for the central thesis of his transcendental idealism in maintaining that space and time are not 'things-in-themselves', but rather 'appearances', or, in other words, that the spatio-temporal character of the objects of possible human experience is grounded in and conditional upon a priori features of our human mode of cognition. According to Kant then "[t]ime is ... merely a subjective condition of our human intuition ... and in itself, outside the subject, is nothing" (A35/B51). Hence, Kant maintains that time is "transcendentally ideal".

[4] Heidegger's interpretation of Kant's theoretical philosophy is presented in his *Kant and the Problem of Metaphysics* (1997[1929]).

It is perhaps in respect of his transcendental idealism, a position concerning the a priori limiting conditions of the objects of possible human knowledge, that Kant's theoretical philosophy is most well-known. However, it is also in discussions of this position that exchanges between Kant's interpreters have tended to be at their most heated, with many commentators finding little to agree upon other than that transcendental idealism involves some kind of distinction between 'appearances' and things-in-themselves, and the restriction of all possible knowledge to the former of these. Contrary to the views of many of his interpreters, however, Kant's denial of the transcendental reality of time need not be read as consigning temporality to a second-class ontological status. This is a point upon which advocates of a 'methodological' interpretation of Kant's idealism, such as Henry Allison, are especially insistent.[5] Whereas Sartre follows a now familiar line of Kant interpretation in criticizing transcendental idealism for its alleged ontological commitment to a superfluous and unknowable thing-in-itself whose existence, although intended to ground the reality of sensible phenomena, in fact demotes them to the rank of mere subjective representations, it is by no means clear that the Critical philosophy recognizes any such demand for supersensible entities in accounting for the possibility of what we know as empirical reality. Against the Sartrean allegation that the transcendental idealism of the Critical philosophy serves merely to multiply our philosophical perplexities by committing us to the existence of an unknowable thing-in-itself which serves as the ontological ground of temporal phenomena, we may therefore reply on Kant's behalf that this is to miss the point of the Critical philosophy's revolutionary critique of the metaphysical rationalist tradition, according to which all discourse concerning things-in-themselves is idle speculation.

The 'idealism' of Kant's position concerning the a priori form of sensible intuition rests not upon any supposed numerical identity between the objects of spatio-temporal experience and our representations of them (a position which Kant rejects as 'empirical idealism') but rather upon a claim about the subjectivity of certain a priori formal conditions of the possibility of reference to experiential phenomena, namely, space and time. In maintaining that space and time are transcendentally ideal, Kant holds that they are not objects in their own right. The representations of space and time do not, according to Kant, refer to anything independent of the subject's a priori form of sensible intuition. Hence, Kant maintains, in representing space and time we are

[5] See especially, Allison (2004).

not representing objects as such, but rather subjective horizons for the representation of objects.

Kant argues for the transcendental ideality of time in both the Aesthetic and the Dialectic, those sections of the first *Critique* which examine the faculties of sensibility and reason, respectively. In the Aesthetic, Kant's argument for the transcendental ideality of time is based on an appeal to the apriority of the representation of time and of certain modes of knowledge which Kant claims are grounded in it.[6] Kant argues for the apriority of the representation of time by noting, first, that all humanly possible experience presupposes a temporal order of succession which cannot therefore be derived from experience, and, second, that, although one can represent to oneself time without empirical content, the reverse is not the case. However, Kant maintains, if it is not by means of sensible experience that we acquire the representation of time, then its relation to empirical phenomena will remain inexplicable unless we hold that temporality is not an ontological condition of objects in general, but, rather, a subjectively originating condition of the possibility of cognitive access to the objects of possible experience. Hence, according to Kant, the temporality of the objects of possible experience is to be explained in terms of their relation to the subjective conditions of an object's being immediately given to the subject, rather than in terms of features inherent in the object itself. To assume otherwise, and to hold that we have non-empirical access to a temporal state of being which obtains irrespective of any relation to our mode of cognition, is merely dogmatic, however.

The transcendental ideality of temporality is further demonstrated, Kant maintains, by exposing the paradoxical implications of treating time as a thing-in-itself. In the Dialectic, Kant discusses a set of Antinomies, each of which is intended to undermine the plausibility of transcendental realism by showing that insurmountable difficulties arise from assuming that what he has presented as the human subject's a priori modes of sensible intuition are instead things-in-themselves. The First Antinomy, in particular, addresses a conflict between two competing metaphysical hypotheses concerning the spatio-temporal bounds of the world, where each hypothesis is meant to be demonstrated by a

[6] In the first *Critique*, Kant refers to certain items of synthetic a priori knowledge which we possess concerning temporality itself, such as that time has only one dimension and that different times are successive rather than simultaneous. In the *Prolegomena*, Kant also emphasises a role for the pure intuition of time in accounting for the possibility of synthetic a priori arithmetical judgements.

proof of the impossibility of the other. According to the Thesis of the First Antinomy, "[t]he world has a beginning in time, and in space it is also enclosed in boundaries", whereas its antithesis holds that "[t]he world has no beginning and no bounds in space, but is infinite with regard to both space and time".[7] The argument for the thesis, so far as it concerns temporality, is that if there were no first moment in time, then an infinite amount of time must have passed in order for the present moment to have been reached. This is impossible, however, since the infinity of a series consists in its being impossible to complete. However, the argument for the antithesis, with respect to temporality, maintains that there can be no first moment in the history of the world because every beginning entails a previous moment in time from which it has emerged and the world could not have spontaneously arisen from a time absent of real content. As such, Kant maintains, transcendental realists have conclusive grounds for subscribing both to the thesis and to the antithesis, and are therefore trapped in contradictory commitments. Transcendental idealism offers us a way out of the Antinomies, however, by allowing us the option of not having to decide between the two competing hypotheses. Because time is not a thing-in-itself, there is no fact about how it is independently of its capacity to yield intuitions for us.

3 Sartre on being-for-itself as the source of temporality

As is readily apparent from his discussion, in *Being and Nothingness*, of the three temporal 'elements' of past, present and future, Sartre is just as unwilling as Kant to attribute temporality to anything taken independently of the subjective conditions of human experience in general, or what he calls the 'being-for-itself' of spontaneous, lived experience, as distinct from the 'being-in-itself' of inert, lifeless matter. Sartre writes, for instance, that "[i]t is through the for-itself that the past arrives in the world" (BN.137), that "the For-itself is the being by which the present enters into the world" (BN.145) and also that "it is only by human reality that the Future arrives in the world" (BN.147). For Sartre then, as for Kant, an account of temporality is an account of a fundamental subjective condition of the reality to which human consciousness addresses itself.

[7] The thesis of the First Antinomy is at A426/B454, whereas the antithesis is at A427/B455.

However, despite their common commitment to the view that human consciousness performs a constitutive grounding role with respect to the temporal reality to which it imparts a synthetic unity, Sartre denies, whereas Kant maintains, that the necessary conditionality of temporality itself upon specifically subjective grounds of its possibility necessarily entails a commitment to some form of idealism. A hostility to all forms of idealism, including Kant's, is in fact apparent throughout *Being and Nothingness* and it is central to the project of that work to develop an alternative to realist as well as idealist positions concerning the relation between consciousness and its intentional objects, since neither approach can, in Sartre's view, properly accommodate the nature of our lived experience as elaborated in phenomenological description. Although he shares Kant's suspicions concerning the possibility of explaining our relation to the phenomena of experience if they are to be taken as things-in-themselves unconditioned by any necessary relation to human subjectivity, Sartre criticizes idealism in general for its narrow conception of the subject-relativity of experience as the relation of an object of cognition to a subjective state of knowledge. Hence, for Sartre, idealism exaggerates the philosophical priority of epistemology by proposing to determine the world in terms of its relation to our capacity for knowledge, rather than in terms of what is implied by the kind of being specific to subjectivity. As Catalano writes, "eventually we must recognise that knowledge must be an aspect of being itself" (Catalano 1974: 31) and hence:

> Sartre's basic objection to idealism is that a philosophy of knowledge must be based on a philosophy of being: that is, the question of how we know reality presupposes an answer to the question of the "nature" of reality. Sartre thus wishes to separate himself from a long tradition – beginning with Descartes and Kant – that claims, in general, that we should first investigate the workings of the mind before examining to what extent the mind knows reality. (Catalano 1974: 31)

Sartre therefore criticizes idealism for its bias in favour of specifically cognitive modes of relation to phenomena at the expense of more basic non-cognitive forms of experiential engagement which are vital to any phenomenologically accurate portrayal of our epistemic predicament as beings actively participant in a surrounding environment which is of practical concern to us. For Sartre, then, idealism distances us from the world in its manifold layers of phenomenological significance by

presenting our principal mode of relation to phenomena as cognitive. In Sartre's view, it is not primarily as a spectator of objects of epistemic interest that consciousness addresses itself to the phenomena of experience, but rather as an agent already preoccupied with circumstances of personal concern So, whereas the object-constitutive activities of the Kantian transcendental subject are geared towards the first-personal representation of an austere domain of naturalistic phenomena subordinate to deterministic physical laws and explicable in terms of the exact sciences, Sartre presents human subjectivity as responsible for the constitution of a world of lived experience whose horizons of significance are no more limited than the spontaneous creative potential we display in our various efforts to interpret our environment from any number of adopted roles and perspectives, among which that of the scientist is but one mode of comportment, and far from being the most fundamental. The influence of Heidegger's rejection of epistemology in favour of 'fundamental ontology' is readily apparent here, and, like Heidegger, Sartre proposes to conduct his ontological investigations by means of a phenomenological account of the being of human reality, because it is we for whom phenomena are experienced as meaningful, and an elaboration of our existential familiarity with the world of lived experience may therefore be expected to assist in the clarification of the conditions under which our situation conveys itself in terms of its significance for consciousness.

Although Sartre departs from Heidegger in attributing greater importance to nothingness than to time in his account of the features of human subjectivity which condition the possibility of an understanding comportment towards being, there can be little doubt of the significance of temporality to Sartre's phenomenological ontology. Unlike Kant, whose pre-eminent theoretical concerns with the possibility of exact science lead him to ground temporality in a subjective faculty the principal function of which is to determine the non-conceptual horizons within which sensible phenomena may be located so as to be addressed in object-directed acts of conceptual thought, Sartre presents temporality as originating in the efforts of human subjectivity to address itself to its own being and that of its lived environment. According to Sartre, pure being-in-itself, the ontological category inhabited by inert, lifeless matter, lacks the capacity for intentionally directed acts of consciousness, so that nothing can become an issue for it and it is therefore comparable to a simple compact substance, without internal differentiation and incapable of distancing itself from its own being in order to determine its relation to it. Being-for-itself, the ontological category inhabited by

human subjectivity, could not be more different, however, for the for-itself is that which is fundamentally self-determining in consequence of its inalienable potential to distance itself from its being and reinterpret its relation to itself and its circumstances. Sartre therefore calls being-for-itself a 'nothingness' because of its ability to spontaneously negate any determinate mode of its being and its consequent lack of any fixed essence. Temporality is, according to Sartre, a phenomenon grounded in this nothingness, arising out of the inalienable condition of the for-itself as a self-determining being and providing certain horizons within which human subjectivity is necessarily conditioned to address and interpret itself. As Catalano puts it, for Sartre, "[t]ime thus comes to being only through a reality that is temporal in its own reality" (Catalano 1974: 111), meaning that it is out of the original self-temporalizing of the for-itself that temporal phenomena result. Again, according to Levy, "[t]he for-itself is the being who separates itself from itself by secreting nothingness, thus consigning what it is to its past. To put it another way, by temporalizing itself the for-itself brings temporality into being" (Levy 2002: 56).

According to Sartre then, an understanding of the temporal structure of lived experience is necessary for any adequate grasp of the for-itself as that being which, in actively distinguishing itself from itself, projects itself from its immediate presence to being, away from its past identity and towards its future possibilities. Hence Sartre writes that:

> It is "in time" that the for-itself has its own possibilities in the mode of "not being"; it is in time that my possibilities appear on the horizon of the world which they make mine If, then, human reality is itself apprehended as temporal, and if the meaning of its transcendence is its temporality, we can not hope to elucidate the being of the for-itself until we have described and determined the significance of the Temporal. Only then shall we be able to approach the study of the problem which concerns us: that of the original relation of consciousness to being. (BN.129)

In maintaining that temporality is the meaning of human reality's 'transcendence', Sartre holds that it is in time that the for-itself overcomes or goes beyond that which is immediately given, and thereby distinguishes itself from the essential self-identity of being-in-itself by projecting itself towards unrealized possibilities of being.

While insisting upon the need for a unified treatment of time as a synthetic unity, Sartre nonetheless thinks it possible and necessary to

distinguish between the temporal elements of past, present and future, each of which displays the for-itself in a different mode of comportment to being. Sartre could perhaps agree then with Bergson's allegation that "Kant's great mistake was to take time as a homogeneous medium" (Bergson 2005[1889]: 232) insofar as Sartre's own ontology recognizes a need for temporality to be treated in terms of what he calls the 'temporal dimensions' of past, present and future. The synthetic unity of these temporal dimensions is, according to Sartre, grounded in that of being-for-itself.

According to Sartre, the past, for example, is in each case 'mine'. That is, the past is always the personal history of a specific being-for-itself which carries its previous experiences with itself and, in so doing, sustains in being that which has gone before. Hence, for Sartre, the past owes its survival, its non-erasure from being, to the continued existence of the particular for-itself whose personal history it is. The relation which being-for-itself has to its past is always somewhat paradoxical, however. On the one hand, being-for-itself is identical with its past. The past is, for each particular for-itself, its own unalterable personal history for which it must take ultimate responsibility. As Catalano suggests, for Sartre, "[t]he past is the human reality as it approaches the in-itself, that is, as it approaches the identity of a thing" (Catalano 1974: 114), for one does not have the freedom to alter one's own history, and one's previous existing therefore solidifies into a self-identical essence. At the same time, however, being-for-itself exists as the constant activity of distinguishing itself from all positive content, including that which is constitutive of the identity which it has crafted for-itself through its previous undertakings and abstentions. As such, the absolute freedom which Sartre attributes to being-for-itself is not compromised by the facticity of its past.

4 Sartre's criticisms of Kant's account of temporality

Sartre raises three main criticisms against Kant's treatment of temporality. First, Sartre objects to Kant's efforts to ground the experience of temporality in the workings of a transcendental, and therefore non-empirical, subject, claiming that the introduction of synthetic unity from a subjectivity external to the empirical domain cannot accommodate the internal relatedness of temporal phenomena. Second, and relatedly, Sartre criticizes Kant's attempts to explain the possibility of alteration over time in terms of a persisting substance which undergoes change, maintaining that substantiality and temporality are

ontologically incompatible, belonging to the dichotomous ontological categories of being-in-itself and being-for-itself respectively. Third, and most importantly to his pre-eminent concerns with the existential freedom of human subjectivity, Sartre complains that Kant's attempts to reconcile the deterministic natural order of temporal phenomena with the noumenal possibility of self-determining practical agency by leaving conceptual space for a non-temporal source of spontaneity, commits him to a contradictory understanding of free will, a condition of being-for-itself which, is impossible in abstraction from our temporality.

The first of Sartre's criticisms features in his discussion of 'static temporality', or what Kant calls 'the order of time', which is the relation of temporal phenomena into an irreversible succession of 'befores' and 'afters'. Having noted that such a temporal series involves both a multiplicity of temporal phenomena distinct from one another in virtue of their separate locations in the serial arrangement at issue and the unity of a common successive order within which different times are related precisely in terms of their being prior or posterior to one another, Sartre is concerned to explain the grounds of the synthetic combination in question, so as to account for the diversity or divisibility of temporality conceived in terms of the static relation of before and after as well as its indivisibility or continuity. According to Sartre, previous attempts to handle this issue have either treated the multiplicity of temporal phenomena as primary and the unity of time as secondary or have adopted the reverse strategy, but neither approach has been able to ground the feature of temporality which it treats as derivative by appeal to the characteristic to which it attributes priority. If we take the first strategy and divide time into indivisible atoms or 'instants' with a view to explaining the unity of temporality in terms of the relations of before and after which obtain between the temporal simples in question, then we shall find ourselves in the awkward predicament of being unable to explain how something entirely simple can extend beyond itself to come into relation to something other than itself. The trouble here is that if we treat the unity of time as a construction out of instants then the indivisibility of such temporal atoms entails that there is no passage of time within them, so that we end up in the impossible position of having somehow to construct a temporal sequence from exclusively non-temporal constituents!

Sartre does not interpret Kant as following the Leibnizian-Bergsonian alternative of beginning with the unity of a temporal continuum from which subsequently to abstract ontologically derivative moments, an approach he regards as incapable of accommodating the reality of the

multiplicity inherent in temporality, but instead regards the Critical philosophy as attempting to ground the synthetic combination of multiple instants by appeal to a transcendental subject in relation to which they admit of unification. Such a strategy cannot succeed in Sartre's view, however, because it requires us to treat the transcendental subject either as temporal or as atemporal, where neither option will deliver the required result. If we take the transcendental subject as temporal, then we are faced again with the problem of explaining the synthetic unity of a sequence in time and now require a further ground for the temporality of the self. If, however, we treat the transcendental subject as atemporal, then it is unclear what there is to gain from introducing a further non-temporal term into the equation.

Sartre's second and third objections arise in the course of his discussion of 'dynamic temporality', or what Kant calls 'the order of time', which is the capacity for the future to become present and, in turn, past, thereby introducing the possibility of genuine alteration or change by permitting temporal phenomena to arise from or out of earlier states, as opposed merely to occupying an adjacent place in the successive order of 'befores' and 'afters'. Here Sartre claims that we are faced with a two-fold problem: "Why does the For-itself undergo that modification of being which makes it become past? And why does a new For-itself arise ex nihilo to become the Present of this Past?" (BN.145). In other words, Sartre's concern is to find an explanation for how experience can accommodate a change of tense. Referring to the argument of the "Refutation of Idealism", Sartre maintains that, according to Kant, "change by itself implies permanence" (BN.145) for the reason that wherever there is alteration there must be something which alters and therefore persists over a period of time over which it undergoes an adjustment in its state. In commenting upon what he takes to be Kant's position, Sartre writes that:

> if we suppose a certain non-temporal permanent which remains across time, temporality is reduced to being no more than the measure and order of change. Without change there is no temporality since time could not get any hold on the permanent and the identical. (BN.145–6)

In other words, for Sartre, Kant's view that change entails permanence of substance commits him to the position that time is conditional upon alteration. This is because that which undergoes change is itself something permanent, so that if we abstract from the alterations which that

substance survives across, we are left with some underlying substratum which is utterly indifferent to temporality. According to Sartre, then, Kant is committed to holding that it is only through the medium of change that time can get any purchase on what is, and since phenomena cannot be represented as temporal except under the condition that they undergo alteration, temporality can only be in operation where there is a process of object-transformation to measure and record. However, since Kant argues in the Second Analogy that "all alterations occur in accordance with the law of the connection of cause and effect" (B232), where such connections are deterministic, it seems to follow that time is inseparable from necessary causation, and the experience of phenomena as temporal is therefore conditional upon the restriction of temporality to an empirical domain of objects governed by natural principles of physical and psychological causal necessity. Here Sartre objects that, although for Kant's position "the *unity* of change and the permanent is necessary for the construction of change as such" (BN.166), the Critical philosophy does not accommodate the required unification of alteration and permanence. This is because, in Sartre's view, Kant's position makes the relation between permanence and change purely 'external' because it rests upon the activities of a transcendental subject combining originally discrete elements.

Turning finally to his third criticism, Sartre addresses "the passages in the Critique where Kant shows that a non-temporal spontaneity is inconceivable but not contradictory" (BN.171) and takes issue with the Critical philosophy on this issue, maintaining that there is in fact a contradiction in the notion of some spontaneous occurrence beyond temporality. As is well-known, Kant's commitment to the position that all alterations in empirical (i.e. spatio-temporal) phenomena take place in accord with deterministic causal laws led him to attempt to ground the possibility of spontaneous human agency in a noumenal domain congenial to the reality of transcendental freedom. According to Sartre, however, spontaneity is itself the necessary ground of temporal experience in that the capacity for original agency entails a projection of oneself away from what is, in an effort to position oneself towards being, where this constant activity of separating oneself from and aligning oneself towards being necessarily entails the rejection of pure being into a past, the announcement of oneself towards a being to which one is present, and the projection of a future towards which one implicitly comports oneself. As such, Sartre maintains that the capacity which the human subject has (or rather is) for spontaneity automatically entails the projection of temporal horizons which spring from the originality

of consciousness in its conscious alignment of itself towards what is. In that case, however, the notion of an atemporal spontaneity is contradictory because it denies that to which it is necessarily committed in the recognition of spontaneous human agency: namely, a projection of our subjectivity away from a being which bears the significance of something past for me, in order to announce myself to my present and towards my future.

5 Conclusion

Whereas for Kant it is our capacity for receptive susceptibility to sensible objects which accounts for the empirical reality of time, Sartre attributes the temporality of human experience to the being of subjectivity as spontaneity. For Sartre, our being is that of an active and creative originality which, in separating itself from the identity of being-in-itself, comports itself towards that being by negating and pushing it into the past, thereby placing itself as present to the world and orienting itself towards its own future. Sartre's criticisms of Kant's views regarding temporality result from a difference between the two philosophers concerning the manner in which temporal horizons enable access to phenomena for human subjectivity, where Sartre is concerned that Kant's pre-eminent interest in explaining the grounds of the possibility of the exact sciences leads him to offer an account of time which, although congenial to mathematics and the physical science of the late 18th century, cannot accommodate any scope for spontaneous human agency, or practical activity not necessitated by deterministic natural laws. Although Sartre is by no means a follower of Bergson, we can therefore see an affinity between his opinion of Kant and Bergson's,[8] in that neither francophone philosopher is willing to accept Kant's principally 'objective' account of time as demarcating a region of interest purely to the mathematical sciences, or at least to the first-personal experience of an empirical domain comprehensible in terms of deterministic physical laws. For Sartre, as for many

[8] In commenting upon the findings of his *Time and Free Will*, Bergson remarks that:

> we have tried to prove that duration, as duration, and motion, as motion, elude the grasp of mathematics: of time everything slips through its fingers but simultaneity, and of movement everything but immobility. This is what the Kantians and even their opponents do not seem to have perceived: in this so-called phenomenal world, which, we are told, is a world cut out for scientific knowledge, all the relations which cannot be translated into simultaneity, i.e. into space, are scientifically unknowable. (Bergson 2005[1889]: 234)

other phenomenologists, Kant's mistake is in privileging the status of mathematical science over other ways in which the empirical domain can be experienced as comprehensible, the natural result of which is that he finds it necessary to introduce a dualism between an empirical domain without spontaneity and a noumenal realm independent of facticity. As has been seen, however, Sartre maintains that it is in consequence of the human capacity for spontaneity that time originally comes into being, so that it is impossible to distinguish between a noumenal and phenomenal self, although the Kantian idealist's commitment to a transcendental as well as an empirical self may make such a distinction seem necessary.

References

Abbreviations used:

A/B: Kant, I. (1998) *Critique of Pure Reason*. Tr. P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Prol: ——(2001) *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science with Kant's Letter to Marcus Herz February 27, 1772*. 2nd ed. Tr. J. W. Ellington. Indianapolis, IN: Hackett.

BN: Sartre, J.-P. (2005) *Being and Nothingness*. Tr. H. E. Barnes. London: Routledge.

Allison, H. (2004) *Kant's Transcendental Idealism: An Interpretation and Defence*. New Haven, CT: Yale University Press.

Bergson, H. (2005[1889]) *Time and Free Will*. Tr. F.L. Pogson. London: Elibron Classics.

Catalano, J. (1974) *A Commentary on Jean-Paul Sartre's Being and Nothingness*. Chicago: The University of Chicago Press.

Cox, G. (2006) *Sartre: A Guide for the Perplexed*. London: Bloomsbury Publishing.

Darnell, M. (2006) *Self in the Theoretical Writings of Sartre and Kant: A Revisionist Study*. New York: Edwin Mellen Press.

Gardner, S. (2009) *Sartre's "Being and Nothingness": A Reader's Guide*. London: Bloomsbury Publishing.

Heidegger, M. (1997[1929]) *Kant and the Problem of Metaphysics*. Bloomington, IN: Indiana University Press.

Levy, N. (2002) *Sartre*. Oxford: One World Publications.

4

Kant and Sartre: Psychology and Metaphysics: The Quiet Power of the Imaginary

Thomas R. Flynn

Foucault once wondered, "What philosopher has not tried to refute Plato", and then he proceeded to add his name to the list. Sartre might well have echoed that thought, substituting 'Kant' for 'Plato' in the process. Having passed what he recalled were some of the happiest years of his life at the École normale supérieure, rue d'Ulm (*ENS*), where he studied philosophy among the neo-Kantians at the Sorbonne, Sartre was clearly imbued with the Critical attitude, even if he spent the rest of his life trying to redefine or escape it. One could say of the Kantian aspect of Sartre's thought what John Locke said of the sense origin of our abstract ideas, that they bore "the tang of the cask they came in". Of course, Sartre's "cask" retained the flavours of Cartesian, Bergsonian, Husserlian, Hegelian, Marxian and many other philosophical elixirs that the bright *normalien* imbibed in the ensuing years, each with its proper trace. While my point is not to uncover an unprincipled eclectic – such was not the case – it serves to underscore the claim that Kant's thought provided an enduring component of Sartre's philosophical thought throughout his life.

As a young teacher, Sartre remarked that he scarcely distinguished between academic psychology and philosophy as such. In her comments on this claim, his adopted daughter, Arlette Elkaïm-Sartre, noted that the two subjects were often taught in tandem in the official syllabi, thus making it easy for the disciplines to overlap.[1] So it was not surprising for

[1] At the time, Sartre was teaching general psychology at a Lycée in Le Havre after his return from a research year in Berlin (1934) "and for a long time after, French school students were introduced to the four classical fields of philosophy: general psychology (later called 'theoretical psychology'), metaphysics, morals and logic. Imagination belonged to the area of psychology" (Elkaïm-Sartre 2004: vi).

him to publish his first three books, *The Imagination*, *Sketch for a Theory of the Emotions*, and *The Imaginary* without feeling the need to emphasize that he was writing "philosophical" psychology and not "empirical" psychology, which certainly figured in these works nonetheless. Indeed, that distinction was most clearly drawn and in greatest detail in the third of these publications.

In what follows, I wish to reflect on several cases (of many) where the presence of Kantian thought is evident in Sartre's work. Sometimes this presence is constructive; other times it seems to block the progress of Sartre's argument. But, either way, Kant's philosophical shadow hovers over these examples of Sartre's philosophical work.

1 Sartre's *Diplôme*

As a prelude to the consideration of Kant's presence in Sartre's thought, let us go to his very first academic work, Sartre's thesis to complete his studies at the *ENS* and qualify to stand for the aggregation exams. His topic for the *Diplôme d'études supérieures* (*DES*) was "*L'Image dans la vie psychologique*" (1927, "The Image in Psychological Life"). Its director was a famous professor of Psychology at the Sorbonne, Henri Delacroix, who subsequently invited Sartre to publish it (minus the tables and other signs of an academic thesis) as *The Imagination* in a series he was editing for Alcan (1936). The fact that Delacroix's series was entitled *Nouvelle encyclopédie philosophique* indicated the intertwining of philosophy and psychology in that era. As Alain Frajoliet notes in his careful study of this still unpublished text: "The Diplôme cuts a difficult path between [the] materialist reductionism and [the] spiritualism [of the thinkers Sartre is studying] in order to attempt to found a *metaphysics of creative spontaneity like the imagination*" (Frajoliet 2008: 74).[2]

With this in mind, it is significant that Sartre discussed approvingly for the most part the concept of a *schème symbolique* in the work of experimental psychologist Auguste Flach (1925).[3] Sartre's use of the "symbolic imagination" à la Flach is discussed in *PP* 444–53. Thanks to Flach and Sartre's probably thorough reading of the three *Critiques* in the Lycées and the École,[4] he respects the role of *Einbildungskraft* (imagination) in

[2] Hereafter *PP*.

[3] See *PP* 391 and 397 as well as Noodlemann (1996: 220–4).

[4] It is reported from the Archive of the Library of the *ENS* that Sartre checked out the works of Kant on a number of occasions.

the first *Critique*, but is evidently more enthused about the role of the Symbolic schema, which is much closer to the aesthetic insights of the third *Critique*. From the very start, Sartre is concerned more about the creative spontaneity of the imagination than by its ability to elicit apt examples for categorial understanding. In the *DES* this amounts to the distinction between "*symbolizing imagination*" and "*schematizing imagination*". In what follows, we shall see that this distinction encapsulates several concepts and terms vital to Sartre's later thought that will appear at crucial moments to help him deal with metaphysical, moral and aesthetic issues that arise from his reaction to problems that occur in the course of his theoretical reflection and practical (ethical, political) concerns.

At the outset, let us simply underscore the presence of two basic concepts that are already at work in the thought of this 22-year-old and will continue throughout the rest of his career. Contrary to popular opinion, Sartre's "dualism" is not the Cartesian duality of mind and matter (*res cogitans* and *res extensa*), as his writings in *Being and Nothingness* have led us to believe. Without denying that Cartesian heritage, his deeper and more lasting dualism is the Bergsonian duality of *spontaneity* and *inertia*, to which he remains loyal in his later works such as the *Critique of Dialectical Reason* and *The Family Idiot,* his multivolume Flaubert study after he had abandoned, or at least set aside, his "philosophy of consciousness".

The second basic thesis, implicit in Sartre's preference for the *schème symbolique*, is his drive to render the individual accessible in its individuality, for which the image is particularly promising. Beauvoir narrates the by now legendary story of Raymond Aron telling his friend that Husserlian phenomenology would enable a person to "philosophize about that cocktail glass in front of them" (de Beauvoir 1966: 162). Though it is unlikely that this was Sartre's introduction to Husserl's thought, as the story implied, his drive towards the concrete, already exhibited in his favouring of the symbolic schema in the *DES*, culminated in Sartre's epiphanic experience of Raymond's description. It had already found literary expression in the title of Jean Wahl's book, *Vers le Concret*, published the year before (1932). Both Beauvoir and Sartre admired that work and mentioned it several times over the years. Consider Sartre's ordering of his masterwork, *Being and Nothingness*, from the abstract to the concrete of existential psychoanalysis, the link between the symbolic scheme and the third *Critique*, Sartre's enthusiastic embrace of Descriptive Phenomenology with its "free imaginative variation" of examples, and his incorporation of Hegelian and Marxian

concepts (the concrete universal) into his subsequent philosophy of history – these, I believe, issue from this deep-rooted drive "towards the concrete".

2 *Transcendence of the Ego* (1937)

The young Sartre seems to have written this bombshell of a work while spending the year at the Maison Française in Berlin (1933–4). It was not published until 1936–7. Of all the works that attract the interest of 'non-Continental' philosophers, this is the most discussed. Its argument is clear, relatively unambiguous and defends a thesis that seems almost 'postmodern' in character.

As its title suggests, it is addressing the Husserlian notion of a "Transcendental Ego", but its target is equally and at this point especially the Kantian notion of the transcendental unity of apperception on which Husserl's concept is based. To start with Sartre's title, the "of" denotes both a subjective and an objective genitive. Sartre is claiming that the transcendental Ego has been "transcended" in the sense that we do not need its services. Specifically, in a wielding of the principle of parsimony (Ockham's Razor), Sartre insists that the unifying and individuating functions ascribed to such consciousness by Kant and Husserl are accounted for by the "temporalizing" character of consciousness itself with its protentions and retentions along with the inertia of what he will later call being-in-itself (as it will develop in *Being and Nothingness*).

The same principle leads him to see that what Kant called the "empirical" ego (the object of our reflective awareness and of scientific investigation) is "transcendent" in the sense that it is "other" than our pre-reflective awareness. If we consider the Ego (the "*I*") as a subject of our actions as in the claim that "I did that", it is distinguished from the ego (the *me*) as object of events in our lives and liable to the interpretation of others. But under both descriptions, Sartre is arguing, the Ego is a "thing amongst things" in the world. That claim plays not only an epistemic and an ontological role in Sartre's thought but a moral one as well.

Sartre begins the second paragraph of the treatise with a deceptively innocent remark: "It must be conceded to Kant that 'the I Think *must be able* to accompany all our representations'" (*TE* 32). Then comes the counterclaim: "But need we then conclude that an *I in fact* inhabits all our states of consciousness and actually effects the supreme synthesis of our experience? This experience would appear to distort the Kantian view," Sartre insists. "The Critical problem being one of validity, Kant

says nothing concerning the actual existence of the *I Think*, [the existential question]. On the contrary, he seems to have seen perfectly well that there are moments of consciousness without the *I*, for he says '*must be able* to accompany' ".[5] In Sartre's initial view, "Kant treated the *de jure* issue but overlooked the *de facto* problem that led some of his followers to appeal to the transcendental consciousness as "a pre-empirical unconscious" (*TE* 33; *TE-F* 1515), a view to which Sartre was opposed even at this early stage.

Like Kant, Sartre was a moralist, concerned about the moral responsibility of individuals and, later in his career, with the systematic distortion of our moral compass by institutions like colonialism. But the moral aspect of *Transcendence of the Ego* does not surface until the last few pages of the work, when Sartre describes the "ego" as a defence against our anguish in the face of spontaneous consciousness. "Perhaps the essential role of the ego," he hypothesizes, "is to mask from consciousness its very spontaneity" (*TE* 100). The individual flees to the certainty and identity of what Sartre will later call "being-in-itself" to escape the anguished creative freedom of pure consciousness. This is the meaning of what, later in *Notebooks for an Ethics* (1992[1947–8]), he will describe as learning to "live without an ego", and which he will describe as a form of "authenticity" often taken for the chief existentialist value (Sartre 1992[1983]: 414, 417).[6] In a remark that conjoins metaphysical, psychological, moral and political concerns, Sartre suggests that several of our "psychoasthenic ailments" (*TE* 99; *TE–F* 83) may well flow from "ontological" anguish – a term implicit here but elaborated in *Being and Nothingness* – with its "vertigo of spontaneity" (*TE* 100). He suspects that such afflictions might be treated if we but had the courage to face up to our egos as the hiding place of our creative freedom. This is an anticipation of that psycho-ontological procedure called "existential psychoanalysis" in *BN* that he will illustrate in detail in his subsequent "biographies".

This double attack on both Kant and Husserl issues in perhaps the most significant claim of this book for Sartre's later thought: the famous notion of an 'egoless' consciousness (Gurwitsch 1979: 287). Towards the end of the book, Sartre summarizes his thesis: "a pure transcendental sphere accessible to phenomenology alone" is a sphere of "absolute

[5] Sartre (1957: 32) and Sartre (1972: 13–14), hereafter *TE* and *TE–F*, respectively. For a careful development of this issue, see *PP* 591–612; and de Coorebyter (2000: 177–217), hereafter *SFP*.

[6] See also Grene (1952).

existence". (TE 95; TE-F 77) "We may therefore formulate our thesis: transcendental consciousness is an impersonal spontaneity". *(TE* 98; *TE–F* 79).

Sartre is preparing the way for a "prereflective *Cogito*" in *BN* that will yield the complete transparency that Husserl desired with the transcendental reduction. But the problem, as Sartre came to discover, is that the theory of intentionality (the "realist instrument" that he never abandoned, according to which all consciousness is "of" an other-than-consciousness) does not sit well with the meaning-giving character of Husserlian consciousness (which seems redolent of idealism) or with the Kantian constitutive character of consciousness, for that matter. Sartre gradually made the decision almost as by default to opt for intentionality and a realist epistemology.

3 *The Imaginary* (1940)

This important study gathers the insights of the previous volumes and funnels them into theses and arguments that will appear for the rest of Sartre's career. It is the bridge to *Being and Nothingness*, no doubt, but its quiet power extends throughout his other works so that his massive Flaubert study can be said to be its "sequel" (Sartre 1977: 119).

Before moving to the Husserlian portion of this study – which nonetheless contains a "critical" dimension to its formulation, let us advance momentarily to the penultimate paragraph of the book where Sartre gestures to his neo-Kantian audience, unfamiliar with talk of intellectual intuitions and immediate experience of the thing-in-itself:

> We can now pose the metaphysical question that has been gradually disclosed by these studies of phenomenological psychology. It can be formulated thus: what are the characteristics that can be attributed to consciousness on the basis of the fact that consciousness is capable of imagining? This question can be taken in the sense of a critical analysis in the form: what must consciousness in general be if it is true that the constitution of an image is always possible? And without doubt, it is in this form that our minds, accustomed to posing philosophical questions in the Kantian perspective, will best understand it. But to tell the truth, the deepest sense of the problem can be grasped only from a phenomenological point of view. (*Imaginary* 179)

Sartre explains that the best answer to this question would perform an eidetic reduction (from fact to essence via the "free imaginative

variation of an example") enabling us rigorously to describe the "essence" of consciousness. This, in fact, is what he did in the first part of the book. But, as if that had not happened, he now concedes: "as the idea of an eidetic intuition is still repugnant to many French readers, I will use an oblique method, which is to say a somewhat more complex method."

Anyone familiar with Sartre's later work, especially his *Search for a Method* (1968[1957]), will recognize this as an anticipation of the "progressive/regressive method" employed for the remainder of his work. Its regressive procedure is "analytic", arguing from the fact to the conditions of its possibility. But, of course, these conditions are rarely "transcendental". Chiefly they would include, for example, the condition of French bourgeois society or, more specifically, the state of poetic activity in France during the second and subsequent quarters of the 19th century (as in the cases of Beaudelaire, Malarmé and Flaubert). But the essence grasped by eidetic reduction, insofar as it is acontextual and timeless, could well be labelled "quasi-transcendental", closer to what Husserl termed belongs to a "regional ontology" or Foucault, salva reverentia, the "historical" a priori.

Returning to the opening chapter of the book, we find an almost textbook application of Husserlian phenomenology to imaging consciousness. After a careful eidetic reduction to reach with "certainty" the essence of the imaginary, Sartre leads us on an ambitious tour of current theories of imaging consciousness in his day. What I wish to emphasize, however, is the enthusiasm with which he claims pride of place for imaging consciousness among our psychic functions (except, grudgingly, perception – he hopes to remain a realist, after all) (Sartre and Verstraeten 1991: 83ff).

Revealing that Sartre has been reading Heidegger's major text in his own manner, he now asks in telescopic fashion:

> Is not the very first condition of the *cogito* doubt, which is to say the constitution of the real as a world at the same time as its nihilation from this same point of view, and does not the reflective grasp of doubt as doubt coincide with the apodictic intuition of freedom? (*Imaginary* 186)

He concludes that imagination is thus not an empirical power added to consciousness but is "the whole of consciousness as it realizes its freedom". In sum, "it is because we are transcendentally free that we can imagine" (*Imaginary* 186). But he reverses the relationship and extends

the claim: "The nihilating function belonging to consciousness – which Heidegger calls *surpassing – can be manifested only in an imaging act*" (*Imaginary* 186–7, emphasis added).

Sartre is marshalling his earlier remarks on nothingness, throughout these three psychological studies and even from his earlier works, to undertake a creative dialogue with the Heidegger of *Being and Time*. Not that Heidegger inspired the idea – we have noted its presence at work even before Sartre's "Berlin vacation" – but that German masterwork certainly challenged an equivalent response, the initial elements of which are sketched in this portion of Sartre's concluding remarks. We glimpse what will be a basic claim of *Being and Nothingness*, namely, that human reality is being-in-situation; that "situation" is an ambiguous relation of facticity (the real world) and transcendence (the surpassing of that real towards the irreal or imaginary).

So the imaginary is that concrete "something" towards which the existent is surpassed. As soon as a person apprehends his or her existence as "in-situation", Sartre is claiming, they surpass it towards that in relation to which the person exists *as lack* – their possibilities: goals, values, "as ifs". But the locus of that lack is the imaginary. In effect, the "imaginary represents at each moment the implicit sense [*sens* (meaning/direction)] of the real" (*Imaginary* 188). We shall return to the Kantian inspiration of those "as ifs" shortly.

Continuing this quasi-apotheosis of the imaginary, Sartre urges that "the object of a negation must be posited as imaginary", adding that "this is true for the logical forms of negation (doubt, restriction, etc.) as for its affective and active forms (prohibition, consciousness of impotence, lack, etc.). ... There can be no realizing consciousness," he assures us, "without imaging consciousness, and vice versa. Thus imagination ... is disclosed as an essential and transcendental condition of consciousness. It is as absurd to conceive of a consciousness that does not imagine as it is to conceive of a consciousness that cannot effect the *cogito*" (*Imaginary* 188). We must conclude *that imaging consciousness is the locus of negativity, possibility and lack* – features that Sartre will attribute to consciousness in general in *Being and Nothingness*. The imagination has reached its high point in Sartre's philosophy. Henceforth, there will be a gradual reduction of its explicit role in his thought until we encounter it *in extremis* in the replay of the imaginary and the real in the life and work of Gustave Flaubert (Flynn 2014: 133–4).

This cannot be emphasized too strongly because it confirms Sartre as a philosopher of the imaginary and sets our expectations for his subsequent works, not only literary but philosophical and political – which

we might describe as his hyperbolic politics, fruit of the political imaginary.

4 An expansion of the imaginary: the "as if"

An implicit but widespread use of the imaginary occurs in Sartre's frequent appeal to the "as if" (*comme si*) to lend a certain counter-factual support to an observation he is in the process of making or an action he is proposing. These plausibility claims are chiefly interpretive in nature, because they urge us to think or act in a certain way (*as if*) that lends a support to or at least qualifies the imaginary dimension of the statement being proposed. To cite an example from his *War Diaries*, Sartre discusses the "as if" with which he assumes a certain inevitable event, namely, an order his platoon has just received to leave for a new position in the war zone, Morsbronn (Alsace):

> Not to *accept* what happens to you. That's too much and not enough. To *assume* it (when you've understood that nothing can happen to you except by your own hand), in other words, to adopt it as one's own, exactly *as if* one had given it oneself by decree, and accepting that responsibility, to make an opportunity for new advances, *as if* that were why one had given it oneself.
>
> This 'as if' is not a lie, but derives from the intolerable human condition, at once *causa sui* and without foundation, so that there is no judge of what happens to it, but all that does happen to it can do so only *by its own hand* and within its responsibility. (Sartre 1984: Notebook 3, 95)[7]

The model of this "as if" is Kant's famous trio of Ideals (ideas) of Pure reason, namely, the existence of God, the freedom of the will, and the immortality of the soul. Though they cannot yield knowledge because there is no corresponding percept that could bring them under a concept, that is, they cannot be known by Kant's strict condition of placing a percept under a concept, still they can be "thought" and they can inspire thought and action that enjoins one to act "as if" these three ideas named objects of cognition. The point is that one "must" act "as if" each of these ideals obtained in the realm of the in-itself, when in fact one knows that such is not the case and cannot be. But they do enable

[7] See Bürger (2007: 75).

the knower to "tie together" the loose ends of its other knowledge into a sense of the whole or complete account. In other words, of relevance to our topic is the fact that Kant does speak of the need for a *focus imaginarius* to guide our actions in these three domains (A644/B672).[8] This, of course, refers to the regulative, not the constitutive, use of these ideas. Thus, Kant asserts that the search for an ultimate explanation of the universe – the drive behind our pursuit of science – must assume a supreme intelligence that it can only "think", but cannot "know". Its use is hypothetical and pragmatic but necessary for our drive towards completeness. As he phrases it in the matter at hand: "We declare, for instance, that the things of the world must be viewed *as if* they received their existence from a highest intelligence" (A671/B699). Sartre's use of the "as if" is much broader than the original Kantian employment. As I said, it included the interpretive and the rhetorical use of our discourse as well as the moral dimension, as we are about to see. But in the case of each thinker, we are witnessing the quiet power of the imaginary.

5 *Existentialism is a Humanism* (1944)

This is an important document in Sartre's philosophical evolution, expanding the scope of existential responsibility from the individual consciousness (monad) to the circle of human beings alive today (whenever "today" happened to be). Though it is the philosophical work by Sartre that most people read if they read anything of his in this genre, this is the only work that he openly regretted having published. The reasons are clear. It was the stenographic recording of a public lecture delivered without notes as the philosophical movement was bursting onto the Parisian scene. Indeed, it was taken by many as the manifesto for this form of thought and manner of life. The particular value of this lecture is that it presents Sartre thinking "on the wing". It is at once creative and inconsistent, a series of very apt insights (*aperçus*), rather than a fully developed argument. If one were to read it in the context of other essays published at about the same time, a more convincing case could be made for the "collective responsibility" he is beginning to ascribe to his auditors.[9] But as it stands, it is a rhetorical tour de force.

---

[8] References to Kant's first *Critique* follow the standard system A/B, with A referring to the pagination of the work in the first edition, whereas B, for the second edition. The translation used is listed in the References section of this chapter.

[9] I undertake such a rational reconstruction of Sartre's argument, supplemented by his other publications from that period, in my *Sartre and Marxist Existentialism* (1984: ch. 3).

When he turns to ethical matters and the concept of moral responsibility for other human beings, he seems immediately to appeal to Kant's "universal legislator" formulation of the Categorical Imperative. He does so without any justification except for the strategic necessity to expand the scope of existentialist moral responsibility beyond the individualist concern apparently adopted in *BN*.[10] Yet Sartre's Kantian formalism is not as pure as in Kant's own formulations. Sartre is not concerned about the intention of the moral agent so much as with the commitment to bring it about that everyone be "free". The intended freedom is not the abstract "freedom that is the definition of man", which Sartre seems to take as self-evident, but the "concrete" freedom in the ethico-political sense of being able creatively to fashion and follow their own values and in the midst of a society of free individuals. Working for concrete freedom implies commitment to work towards a specific outcome, namely, what Sartre calls "the city of ends" (overthrowing Kant's Kingdom with a word) but later describes as "the human realm", or simply the "ethical". We should note that even in this Kantian formula, Sartre introduces an existentialist dimension, namely, the "anguish" of the free agent. He appeals to what the commander of a regiment must feel as he orders a sortie that he knows will be lethal for half his troops. This "existentialist" aspect of the principle of universalizability reveals that even in the most antiseptic formalism, Sartre is deeply aware of the "existing" individual and the burden that his or her freedom imposes. In this sense, Sartre seems closer to the third formula of Kant's Categorical Imperative ("The Universal Lawgiver in a Kingdom of Ends") than to its sheer "universalization" formula.

Amongst the common criticism of Kantian "formalism" in ethics, Sartre enlists the standard objection of the moral dilemma. In its extreme, one could say "tragic", form, there is no escape without incurring "Dirty Hands", the title of his play that exhibits this dilemma between the competing demands of ethics and politics. The moral imagination is obviously at work here, and the dilemma portrayed can be seen as a kind of "eidetic reduction".

We must contrast this well-known "Kantian" formalist argument with the next, unjustifiably neglected, thesis derived from Max Scheler's

[10] Sartre once described *Being and Nothingness* as an "eidetic of bad faith". In other words, it was a phenomenological "description" of the conditions that foster inauthenticity in our current society. Another study, he promised, would address the issue of an ethics of authenticity – see Flynn (2014: 189 n48) and *BN*: 70 n9.

study of the "material a priori" or what is translated as *Formalism in Ethics and Non-Formal Ethics of Values* (1973[1913–16]). Amid a positive critique of Kantian formal a priori, what interests Sartre in this book is Scheler's introduction of the moral imagination into the discussion. Sartre is now playing Scheler against Kant as the material to the formal a priori. Note the appeal to image rather than to rule in the following argument from *EH*:

> When we say that man chooses for himself, not only do we mean that each of us must chose himself, but also that in choosing himself, he is choosing for all men. In fact, in creating the man each of us wills ourselves to be, there is not a single one of our actions that does not at the same time create an *image* of man as we think he ought to be. (Sartre 2007: 24 – hereafter *EH*; 1970: 25)

As befits a quasi-extemporaneous talk, one geared to promote and defend the values and practices of this contemporary mode of thought, Sartre is not bothered with the articulation of arguments; indeed his "axiological" tendencies at this point seem to militate against such "analytic" case-building. What we might call the "hurried" immediacy of this lecture – which called for a more reflective exchange with several listeners afterward – is doubtless another reason for Sartre's reluctance to publish it in that form.

6 The Gramsci Lecture of 1964[11]

Sartre delivered this lecture at the Gramsci Institute in Rome on May 23, 1964, as part of a symposium on "Ethics and History". It, along with a set of notes for another lecture scheduled for Cornell University the following year and the numerous pages of notes for both, constitute what has come to be known as Sartre's Second, Dialectical Ethics. Again I wish to underscore the difference that Sartre draws between his "ethics" and, chiefly, Kant's Categorical Imperative. As before, it is mainly the universalizing formula that attracts his attention, though the primacy of "duty", which Sartre rather unceremoniously calls the "ethic of a slave", is clearly noted.

Unlike Kant, Sartre did not produce an ethical theory; rather, he produced two and possibly three "sketches for an Ethical theory", just as

[11] For a more extensive treatment of Sartre's "second ethics", see my *Sartre* (2014: ch. 14).

he wrote a *Sketch for a Theory of the Emotions*. Though ethical concerns had shadowed his thought for most of his professional life, it is commonly accepted that Sartre wrote a "dialectical" ethics using the vocabulary and dialectical nominalism of the *Critique of Dialectical Reason*. This second attempt focuses on the image of "integral man" rather than on the "authentic" agent of *BN*. These are ideals, and the ethical imagination is fully engaged. But of interest to the Kantian presence is the contrast that Sartre now draws between the security of an ethic of principles (Kant), which he considers inauthentic, and the risk and anguish that a creative ethics entails. In fact, it seems that risk and concomitant anguish are emerging as signs of an authentic ethical stand for Sartre. In these lectures he draws an almost Bergsonian contrast between an "open" ethics and the "closed" variety. The inauthentic, closed ethics is either hampered by its past and traditional rules or forecloses the extent of our future possibilities.

The concept of norm figures prominently in these lectures. It is the ambiguity of norms, Sartre believes, that accounts for the ambiguous or paradoxical character of ethical judgments. They can generate authentic or inauthentic moralities according as they aim for moral autonomy and the maximization of possibilities opening up for the agent a "pure future" (a term Sartre adopts from de Beauvoir) (de Beauvoir 1975 [1946]: 82).

One of his examples is the dilemma of the Belgian mothers who had taken the medication thalidomide for morning sickness only to discover it producing seriously deformed babies. Curiously, he doesn't take a stand on one side or the other regarding infanticide. His point is not unlike the dilemma faced by the young man in *EH* whether to leave his mother and join the Free French or stay with her. Equivalently, he seems to be counselling here as there: "choose, that is, invent", but do so embracing the risk and anguish such a creative choice entails.

This discussion is held in the context of a critique of Kantian rules and duties, though Sartre is equally critical of the positivists. He revives terms from *BN* like "facticity" to introduce a Heideggerian concept that has been dear to him since the late 1930s, "Historialization". In Sartre's use, it denotes the commitment of an agent to their present situation – the admission of their facticity – in order to move beyond it in creative freedom or to remain the same in repetition (as in inauthentic moralities). Sartre is far from being an ethical naturalist. He shares with Kant the ideal of the autonomy of the ethical (what he sometimes calls the "human" realm in contrast with the "subhuman" realm of victims of colonial or economic oppression and exploitation). But, of course, Sartre

has always harboured an anarchist strain – "libertarian socialist" being the received term – though he is quick to deny it when the question arises.[12]

If, by way of conclusion, we contrast Sartre's light, imaginative sketches fleshed out with works of art of the highest quality such that he was awarded the Nobel Prize for Literature[13] with Kant's robust and self-assured architectonic fashioned to foster Enlightenment values, one is led to acknowledge Sartre's growing sense of the necessary but limited promissory role of the imaginary in the face of the violent power of military praxis. As his friend André Gorz remarked apropos his brilliant "biography" of poet and playwright Jean Genet: "One of the conclusions of the work [Sartre's *Saint Genet*] is that the most radical and intense work of liberation can be effected only in the imaginary, for want of the power to suppress the original condition of total alienation."[14] Perhaps Sartre reveals the moral of the story of his life at the conclusion of his autobiography, *Words*, when he confesses: "For a long time I took my pen for a sword; I now know we're powerless" (Sartre 1964: 159). Well, not entirely powerless, as the effect of his subsequent writings attest.

References

Abbreviations used:

PP: Frajoliet, A. (2008) *Le Première philosophie de Sartre*. Paris: Honoré Champion.

SFP: de Coorebyter, V. (2000) *Sartre face à la phénomenologie. Autour de "l'Intentionalité" et de "La transcendence de l'Ego"*. Brussels: Ousia.

DES: Sartre, J.-P. (1927) *L'Image dans la vie psychologique*. Diplôme d'études supérieures.

BN:—— (1956) *Being and Nothingness*. Tr. H. E. Barnes. New York: The Philosophical Library.

TE: —— (1957) *The Transcendence of the Ego. An Existentialist Theory of Consciousness*. Tr. F. Williams and R. Kirkpatrick. New York: Noonday Press.

TE-F: —— (1972) *La Trenscendence de l'Ego*. Introd. S. le Bon. Paris: Vrin.

Imaginary: —— (2004) *The Imaginary*. Tr. with Philosophical Introd. J. Webber. London: Routledge.

EH: —— (2007) *Existentialism is a Humanism*, including a Commentary on *The Stranger*. Tr. C. Macomber. New Haven: Yale University Press.

Bürger, P. (2007) *Sartre. Eine Philosophie des Als-ob*. Frankfurt am Main: Shurkamp.

[12] I have discussed that anarchist strain in Sartre's thought in "An End to Authority: Epistemology and Politics in the Later Sartre" (1977).

[13] Which he declined in 1962.

[14] Gorz, quoting *Saint Genet* in *Le Socialisme Difficile* (1967: 209 n).

de Beauvoir, S. (1966) *The Prime of Life*. Tr. P. Green. New York: World Publishing/ Lancer Books.
—— (1975[1946]) *The Ethics of Ambiguity*. Tr. B. Frechtman. Seacaucus. NJ: The Citadel Press.
Elkaïm-Sartre, A. (2004) "Introduction to Jean-Paul Sartre's *The Imaginary*", in *The Imaginary*. Tr. with Philosophical Introd. J. Webber. London: Routledge.
Flach, A. (1925) "Ueber symbolische Schemata in produktiven Denkprozesse", *Archiv fuer die gesamte Psychologie,* 52: 369–440.
Flynn, T. R. (1977) "An End to Authority: Epistemology and Politics in the Later Sartre", *Man and World,* 10: 448–65.
—— (1984) *Sartre and Marxist Existentialism. The Test Case of Collective Responsibility*. Chicago: University of Chicago Press.
—— (2014) *Sartre. A Philosophical Biography*. Cambridge: Cambridge University Press.
Gorz, A. (1967) *Le Socialisme Difficile*. Paris: Seuil.
Grene, M. (1952) "Authenticity: An Existentialist Virtue", *Ethics,* 62(4): 266–74.
Gurwitsch, A. (1979) "A Non-egological Conception of a Consciousness", in *Studies in Phenomenology and in Psychology*. Evanston, IL: Northwestern.
Kant, I. (2007[1781/1787]) *Critique of Pure Reason*. Tr. N. Kemp Smith. London: Palgrave Macmillan.
Noodlemann, F. (1996) *L'Incarnation imaginaire*. Paris: L'Harmattan.
Sartre, J.-P. (1936) *L'Imagination*. Paris: Félix Alcan.
—— (1992[1947-8]) *Notebooks for an Ethics*. Tr. D. Pellauer. Chicago: Chicago University Press.
—— (1968[1957]) *Search for a Method*. Tr. H. E. Barnes. New York: Vintage Books.
—— (1964) *The Words*. Tr. B. Frechtman. Greenwich, CT: Fawcet.
—— (1970) *L'Existentialisme est un humanism*. Paris: Nagel.
—— (1977) *Life/Situations. Essays Written and Spoken*. Tr. P. Auster and L. Davis. New York: Pantheon.
—— (1984) *The War Diaries. November 1939–March 1940*. Tr. Q. Hoare. New York: Pantheon Books.
—— (1992[1983]) *Notebooks for an Ethics*. Tr. D. Pellauer. Chicago: University of Chicago Press.
Sartre, J.-P. and Verstraeten, P. (1991) "'I Am No Longer a Realist': An Interview with Jean-Paul Sartre", in *Sartre Alive*. Ed. R. Aronson and A. van den Hoven. Detroit: Wayne State University Press.
Scheler, M. (1973[1913-16]) *Formalism in Ethics and Non-Formal Ethics of Values: A new attempt toward the foundation of an ethical personalism*. Tr. M. S. Frings and R. L. Funk. Evanston, IL: Northwestern University Press.
Wahl, J. (1932) *Vers le concret*. Paris: Vrin.

5

Drawing on Sartre's Ontology to Interpret Kant's Notion of Freedom

Christian Onof

Kant and Sartre's philosophies have both been characterized as centred upon strong notions of freedom. Both thinkers understand freedom as involving the capacity to have done otherwise. These thinkers' metaphysical accounts of the possibility of freedom are, however, very distinct. While, in *Being and Nothingness*, Sartre dismisses any form of determinism out of hand, Kant attempts to reconcile transcendental freedom with causal determinism. As a result, many commentators have viewed Kant as attempting to carry out an impossible task. There are indeed important problems with Kant's solution, and the first aim of this chapter is to focus upon ontological issues that arise from an attempt to make sense of Kant's solution. A second aim will be to show how Sartre's approach to ontology and the nature of consciousness can provide us with ways of addressing Kant's problems. In so doing, I shall not be claiming that the proposed solution represents Kant's views, but showing that it accords with a number of his claims. The chapter is structured as follows. After a sketch of Kant's solution to the Third Antinomy, the cogency of the solution will be seen to rely upon a close link between theoretical and practical spontaneity which has echoes in Sartre. An ontological issue is, however, raised by this solution for which the Sartrean dichotomy of For-Itself and In-Itself[1] provides useful insights. A further problem raised by Kant in the second *Critique* is addressed by critically drawing upon

[1] I shall capitalize Sartre's In-Itself and For-Itself, i.e. the same terminology as in the Barnes translation of *Being and Nothingness*, but with capital first letters, to avoid confusion with Kant's notion of in-itself, for which lower case letters are used. I refer to Kant's first and second Critiques as CPR and CPrR, respectively, the *Groundwork of the Metaphysics of Morals* as GMM and the *Metaphysics of Morals* as MM. Sartre's *Being and Nothingness* is referred to as BN. The translations used for both authors are in the reference list.

Sartre's interpretation of consciousness as the source of negation. The chapter concludes with the outline of a proposal for a form of modal realism in the interpretation of Transcendental Idealism which addresses the problems discussed in the chapter.

1 Freedom in the *Critique of Pure Reason*

Kant's solution to the freedom/determinism problem

In the Third Antinomy, Kant considers the claim that there is a causality distinct from that of nature that is manifested in nature, namely the causality of (transcendental) freedom (A444–5/B472).[2] The thesis defends the claim while the antithesis rejects it. Kant's critical solution of this dynamical antinomy involves showing that both thesis and antithesis can be true insofar as they refer respectively to the intelligible and empirical worlds (A535–7/B563–5). That is, it draws upon a distinction characteristic of Kant's Transcendental Idealism. The perspective of empirical objects is not that of things as they are in themselves: it is possible to think of the way things are independently of how they appear in my experience, and this defines an intelligible perspective.

The distinction between these two perspectives is the key for Kant to the possibility that an event should thus be dually determined. As appearance, it falls under the causal laws characterizing the empirical perspective, but it has an intelligible dimension which leaves open the possibility that from this point of view, the event should be ascribable to our freedom. Kant expresses the matter in terms of the notion of character (A538–9/B566–7). While our actions, viewed as empirical events, can be accounted for causally in terms of an empirical character which contains the subjective principles of action characterizing an individual, they can be viewed as the effect of an intelligible causality, the intelligible character, and as such they are imputable.

Much as the duality of empirical and intelligible perspectives upon the same event seemingly creates a conceptual space for a non-natural causality, this does not tell us how the two stories, i.e. the empirical and intelligible accounts, can be simultaneously true. The problem is one of potential over-determination. For, while it is true that an event can be

[2] References to CPR are given as An/Bm where n and m are the page numbers in the 1781 and 1787 editions, respectively. AA refers to the Akademie Ausgabe: *Kants gesammelte Schriften*. Herausgegeben von der Königlich Preußischen Akademie der Wissenschaften, Berlin 1900–" of which volume VI is *Religion within the Bounds of Mere Reason* and volume XVIII is *Reflexionen*.

viewed from two different perspectives, these are not independent of one another. The intelligible perspective is directed to things as they are in themselves, which, as such, are the ground of appearances. So, if we consider some action in the empirical domain, while it is described causally using the empirical character, it is also grounded in reality in itself, and therefore in the intelligible character.

Kant avoids any over-determination at the level of actions as empirical events by having the intelligible character ground the empirical character (A546/B574). So it is not the case that there is anything else but the empirical character determining the nature of a particular event, but rather the empirical character is to be viewed as itself an appearance in the empirical domain of the intelligible character.

This solution does not, however, obviously do away with the problem of an over-determination of action, but arguably shifts it to that of an over-determination of the agent's empirical causality. For, while the empirical character is grounded in the intelligible character, it is also causally impacted by events in the empirical world. Its development can therefore be given a causal account.

In terms of typical naturalistic accounts of action (e.g. Davidson 1963), what we are considering here is something like a causal account of the beliefs and desires that characterize an individual. So, while a belief–desire model of action could indeed be given a non-naturalistic grounding in an intelligible character, it would remain the case that these desires can also be explained causally in terms of agents' past experiences and/or her physiological make up.

Further filling in Kant's account

Kant does not, in fact, say much about the problem of over-determination of action, aside from pointing out that practical freedom requires that for every event, "its cause in appearance was thus not so determining" that there is no room for a role for a causality of freedom (A534/B562). And in a Reflexion also, he refers to this role of "complement of sufficiency" provided by an appeal to transcendental freedom (Reflexion 5611, AAXVIII: 252.14–15). But where exactly is there any incompleteness? In the Third Antinomy, both thesis and antithesis are based upon causal regresses arising from the fact that a totality of conditions is never provided for any given event by identifying a cause. Indeed, as Kant puts it, for every cause, we must seek a cause for the "causality of the cause" (A444/B472). What he means by that is that for any cause, it was not always operational, and therefore, there must be a cause of its causal power becoming effective. When considering an action as the event

to be explained causally, we can thus appeal to an empirical motive as causally responsible for it, and this motive's causal force can be traced back to the agent's past and/or environmental factors.

But there are always many possible options for such regresses. This is because, when seeking causes, one assumes *given circumstances*, but there are different ways of defining what constitutes the background circumstances. If a rock rolls down onto a frozen lake, the subsequent event of its falling to the bottom of the lake when the ice thaws can be viewed as resulting from the thawing taking place under the rock, which itself is caused by the sun's heating. Or it can be viewed as the effect of the rock's having rolled down onto a lake about to thaw, an event triggered by some event at the top of the mountain. When considering an alternative to a given regress, we are identifying a complement of determination of the cause of the original regress. Aside from the thawing which precipitated the rock into the depths of the lake, another cause was required to bring the rock onto its surface in the first place.

Similarly, when considering an action, there are typically two types of causal accounts which can be given (A554–5/B582–3). We can examine the causes of the agent having an empirical motive to carry out a certain action as lying in the agent's past and/or environmental factors. But, when *judging* the agent's action, we view this action as the result of the agent's decision. That is we consider "that this deed could be regarded as entirely unconditioned in regard to the previous state" (A555/B583). On Kant's account, this amounts to introducing a role for the agent's intelligible character. That is, given the environmental conditions and the agent's past, there is still more that is required to account for the empirical character's causality as it is instantiated in this particular action. This is where the complement of sufficiency would lie in the case of human actions.

It is questionable whether Kant has in fact done enough to motivate an appeal to an intelligible character. While it is indeed correct that there are different possible paths that a causal regress can follow, it is not clear why one of these paths could take us outside the domain of appearances to the intelligible character. Can the appeal to the way we judge agents as responsible for their acts provide us with a justification for what is apparently a metaphysical claim about a non-natural causality?

Interpretative options

One way of answering this question in the affirmative will rely upon deflating the status of the claim from a metaphysical to a methodological one. That is, the appeal to an intelligible point of view is not to be thought of as involving a metaphysical claim about how things are

in themselves independently of any particular perspective, but rather to amount to a different perspective where things are considered independently of our knowledge of them (Allison 2004: 16). On such an interpretation, when Kant refers to the intelligible character, he is not adducing some metaphysical reality that lies beyond the world of appearances, but he is referring to a different way of considering agents, namely a way of viewing them independently of their embeddedness within a causally deterministic natural world (Allison 1990: 46). The fact that the natural world, for Transcendental Idealism, is not a world of things as they are in themselves allows us to take a different perspective upon actions, such as one which takes agents as responsible for their acts.

Although this may seem insufficient to motivate an appeal to the intelligible character, we must recall that Kant's aim in the Third Antinomy is 'merely' to show that a non-natural causality as instantiated in human agents is not an impossibility. To show that there is the conceptual space that is required for something like an intelligible character is all that is required. I think that this interpretation achieves this aim, and it must be seen as a strength of the methodological interpretation of Transcendental Idealism.

For reasons which cannot be examined here, the methodological dual-aspect interpretation of Transcendental Idealism which allows for such a non-metaphysical account of the Third Antinomy is problematic in many ways, and I shall not consider its deflationary approach to provide a satisfactory interpretation of Kant's claims about the non-impossibility of a non-natural causality of reason. By making this interpretative choice, however, we are left with a lack of guidance from Kant's text in the solution to the Third Antinomy as to exactly why there is conceptual space for a non-natural causal regress. To repeat, the problem is that it is clear that there is room for more than one causal regress to explain the empirical character's contribution to a particular action, but why should we consider that an intelligible character might be involved?

Again, one might want to recall the modesty of Kant's claim which is just about the non-impossibility of such an involvement. Whether or not this is sufficient to address the problem would seem to hinge upon the nature of the possibility which is at stake here. Indeed, if logical possibility is the notion Kant is referring to, then the fact that more than one causal regress can be adduced to account for the empirical character's bringing about a given action is indeed sufficient to justify the non-impossibility of a non-natural causal explanation, as long as non-natural causality is viewed as free from logical contradiction. Insofar as non-natural causality is not temporal, it is often taken to be problematic.

However, the category of causality in its non-schematized form, i.e. the relation of ground and consequence, can be appealed to in thinking about the intelligible without any contradiction thereby arising. The logical possibility option is therefore attractive insofar as it makes the modesty of Kant's aim in the Third Antinomy sufficient to provide a solution to our problem.

But is it logical possibility that Kant has in mind when he reminds us that all that he is seeking to establish is the non-impossibility of freedom as a non-natural causality? I would argue that it cannot be that, since the category of possibility-impossibility (the first modal category) is applied when there is agreement with the formal conditions of experience (A218/B265–6). This amounts to a stronger notion of possibility, namely one which refers to transcendental logic. Since Kant does not indicate that he is referring to any different (e.g. weaker) notion of possibility in the Third Antinomy, we have no reason to interpret the notion of possibility as a purely logical one.[3]

Piecing together an account

These interpretative choices leave us with little help from Kant's actual text to show the non-impossibility of a role for a causality of reason. To fill in the account, it is useful to consider the agent's situation at time *t*, with an empirical character which has been evolving over time. The conscious agent perceives his environment at *t*, and takes a certain course of action, apparently in response to this perception. It is the empirical character, on the empirical account of action, whose causality is effective in bringing about this course of action.

The empirical character is defined by subjective principles of action characterizing the agent's behaviour in different situations. Does the agent's past, together with the environmental factors, entirely determine how the agent will act at *t*? The problem of free will/determinism arises because it seems that there are two possibilities:

(i) The situation encountered is seemingly identical to a past one, so that the response is determined to be the same, one defined by the empirical character's subjective principles.

[3] One might think that Kant's referring to non-impossibility is a way in which he actually wants to weaken the notion of possibility at stake. Indeed, one might wonder why non-impossibility is not the same as possibility. I think it is reasonable to interpret the difference between the two as the difference between a speculative outline of what might be the case given the conditions of experience, and a construction of what might be the case, as might be carried out in mathematics.

(ii) The situation encountered is novel and will require a further determination of the causality of the empirical character, i.e. further principles specifying how to act in this situation.

While case (i) is the most troublesome for any account that appeals to a role for a non-natural causality, case (ii) also leaves it unclear why an appeal to something outside natural causality may be warranted.

In fact, there are good grounds for viewing the further specification of the empirical character's nature as not completely causally determined by its past together with the given circumstances. This follows from the fact that to decide how to respond in the situation which presents itself at time *t*, that situation must first be assessed by cognizing it, and this is a process involving the agent's spontaneity. That is, that the situation is identical or not to a past situation is only relevant insofar as the agent judges it to be identical or not. And when it is judged not to be, a further judgement is required in which the agent will choose a general principle to apply here (whereby that principle may be the pursuit of some inclination, or a moral principle). Such judgements are, of course, not constrained by anything outside reason. Practically, this means for instance that it is an open matter how the agent will conceptualize the situation in terms of the kind of response that is appropriate for it.

So, for instance if I see someone in need of help but decide to ignore him, I will be conceptualizing the situation as a disturbance in my environment that is to be ignored for the sake of the pursuit of my ends, rather than as a situation that calls upon me to provide assistance of some sort.

By thus appealing to the theoretical spontaneity of reason, I am not attempting to derive transcendental freedom from it. This is a strategy that Kant chose in GMM III for instance (GMM: 448), but which he did not pursue in the second *Critique*. Rather, this appeal serves the purposes of addressing problems (a) and (b). With case (a), it shows that there is conceptual space for a complement of determinacy of my action in the form of a conceptualization of the situation, even though this situation may then be deemed sufficiently similar to a past one for which my empirical character has appropriate maxims defining how to respond. With case (b), such a conceptual space can additionally be filled by appealing to a non-natural ground which would provide an appropriate principle applicable to this situation. Thus the close parallel between practical and theoretical spontaneity (see also Allison 1990) serves as a ground for conceiving of a possible non-natural causality by filling in

the account of the Third Antinomy through the role played by practical deliberation in making a free choice of the ground of action.

Parallels with Sartre

It is interesting to consider the similarities between this interpretation and Sartre's strong notion of freedom. An important feature of Sartre's conception of freedom is that it is not a property of the beings that we are. Rather, it is what we are, i.e. it amounts to a characterization of what it is to be the kinds of being that we are (BN: 25).

Indeed, this follows from the fact that Sartre understands our kind of being as fundamentally distinct from that of things, i.e. of what Sartre calls the In-Itself (BN: xli). This is the being of the For-Itself, which has no place in the fullness of being of the In-Itself, but is essentially a power of negation (BN:11); Sartre's ontology understands the origin of this power as lying in the nothingness that is at the heart of our being as conscious entities (BN: 23). There is a pre-reflective level of consciousness which, as all consciousness, is self-conscious (BN: xxviii).[4] But in this self-consciousness, there is no coincidence between that of which I am self-conscious, and this 'I' which is self-conscious: the nature of self-consciousness is such that there is 'nothing' in this gap between the two relata of the relation of self-consciousness (BN: 78). There is no being In-Itself separating the two relata, but there is nothingness (see Onof 2013).

This same nothingness which separates the relata of pre-reflective self-consciousness also relates my being at any time to my being at a prior time which translates in terms of my freedom to act independently of what has gone before (BN: 27).[5] Sartre does not only reject determinism, but explicitly associates the claim that one's actions are determined (BN: 33) with a form of bad faith.[6] In effect, this follows from a translation into the Sartrean framework of the claim that if our actions were determined, we would not be free to act, but with the additional twist that any claim that there is indeed such determinism, is motivated by a project of attempting to evade the responsibility we have as free beings.

---

[4] As Baiasu (2011) explains, this notion is closely related to Kant's unity of apperception, but there are important differences insofar as the latter involves conceptual activity.

[5] As Gardner (2009: 150) points out, 'the same freedom-constituting cleavage is encountered in the synchronic structure of motivation', referring to Sartre (BN: 34).

[6] Sartre describes those who engage in this form of behaviour as "les salauds" ("the bastards") (Sartre 1970).

There is no space here to carry out a proper comparison of the Sartrean and Kantian conceptions of freedom (for in-depth analyses, see Gardner 2008; Baiasu 2011). What is important for our purposes is:

(i) To note that Sartre's conception of freedom is essentially that of a spontaneity that has the freedom of indifference that is also required for instance, according to Kant, for the blameworthiness of immoral actions (A555/B583). That is, freedom for Sartre is essentially the freedom to do/have done otherwise. Sartre's condemnation of those who claim that we are determined, and would therefore have to account for our notion of freedom in compatibilistic terms is also echoed in Kant's complaint about the dominant compatibilism of the Leibnizian school: "it is a wretched subterfuge to seek an escape in the supposition that the kind of determining grounds of his [(the agent's)] causality according to natural law agrees with a comparative concept of freedom" (CPrR: 95–6).
(ii) To note that Sartre does not, in effect, differentiate between theoretical and practical freedom. This identification is stronger than the link we have exhibited in Kant's accounts of theoretical and practical spontaneity, but suggests some further degree of compatibility between the Sartrean notion of freedom and Kant's notion of transcendental freedom as I have interpreted it.

Both points will be useful in our further investigation of how Sartre's understanding of freedom can be useful to address problems with Kant's.

2 Kant's notion of transcendental freedom in the second *Critique*

There are, of course, further interpretative problems associated with Kant's notion of transcendental freedom in the first *Critique*. Because of the focus of this chapter on how Sartre's ontology can be of assistance in making sense of Kant's understanding of transcendental freedom, I would like to turn to the much less discussed section 'Critical Elucidation of the Analytic of Pure Practical Reason' in the second *Critique*. In this passage, Kant starts with a recapitulation of the outcome of the discussion of the Third Antinomy by reminding us of the conclusions of what he presented there as speculations about the possibility of there being an intelligible character which would ground an agent's empirical character (CPrR: 98–100). In the second *Critique*, with the moral law grounded

(CPrR: 47), this is no longer the speculation which Kant presented in the first *Critique* and which was intended to show that there is a conceptual space for the notion of transcendental freedom in a transcendentally idealistic framework.[7] The fact of reason having grounded the universal normativity of the moral law, transcendental freedom as a condition for the possibility of doing one's duty, is thereby given a practical foundation (CPrR: 47). And as a result, the notion of intelligible character is now "thickened", as Allison (1990: 140) points out: it is now specifically connected with the moral nature of the agent. Thus, Kant indicates that, "if we were capable of an intellectual intuition of the same subject" (CPrR: 99), we would find that "the entire chain of appearances, with reference to that which concerns only the moral law, depends upon the spontaneity of the subject as a thing in itself" (CPrR: 99). Aside from the change of modality to an assertoric mode that is only tempered by its epistemologically conditional status, such a statement would also seem to involve a characterization of the subject's spontaneity (presumably practical, i.e. our capacity for transcendental freedom) as a feature of the subject as thing-in-itself.

This identification is problematic for a number of reasons, of which I shall examine two. First, it is important to recall that the account of the non-inconceivability of transcendental freedom which Kant gives in the Third Antinomy does not, strictly speaking, involve identifying our capacity for transcendental freedom as a feature of our noumenal selves. Rather, Kant claims that, in light of the fact that we have theoretical spontaneity, we can view ourselves as not entirely members of the world of appearances (A546–7/B574–5). This certainly hints at our participation in what one might want to call the intelligible world, and according to Kant's understanding of the concept of 'noumenon', this can be described as a noumenon in the negative sense, i.e. some "thing insofar as it is not an object of our sensible intuition" (B307). Kant can then make a further

[7] Kant clearly states that he has not tried to establish the reality or even the possibility of freedom (A557–8/B585–6). This is because there cannot be any knowledge of the realm of reality as it is in itself. What Kant is showing is that the possibility that reason has causality does not lead to any contradiction with the deterministic order of appearances (A548–9/B576–7, A551/B579), where 'contradiction' must refer to transcendental logic. This in effect amounts to showing the possibility in the world of appearances of some effect of a causality of reason, while of course not showing anything about reality in-itself. Consequently, one finds Kant claiming the possibility of some effect of freedom among appearances (A544/B572), which does not contradict his overall claim that he is not showing the possibility of transcendental freedom.

move to a positive notion of noumenon, i.e. an object of a non-sensible intuition (B307), because in the second *Critique* he has found grounds, namely moral ones, for viewing moral agents to be endowed with transcendental freedom which, according to the Third Antinomy, can only be understood as a feature of the intelligible world.

Noumenal selves as substances?

Standardly, this conclusion is taken to imply our belonging as substances to the realm of things-in-themselves. But this is to make a move from an epistemological to a metaphysical claim. Much as Kant himself is not consistent in his use of the distinctions phenomena/noumena and appearances/things-in-themselves, they are different distinctions (see Allison 2004: 57–9). Mostly, this is not problematic: if something is understood as an object of an intellectual intuition, it is thereby understood as existing in-itself. And if something exists in-itself, an intellectual intuition could intuit it as it is, i.e. in-itself. The problem here is that all we have is a power of choice, transcendental freedom: there is no claim about a substance.

While it is true in the world of sense that causality cannot be understood in separation from substance insofar as the second Analogy defines change in terms of alteration of a substance (B233), when it comes to using the category of causality beyond appearances, this no longer holds. It is only insofar as events in the world of appearances are defined as changes, i.e. in essentially temporal terms, that a substrate of change is required, i.e. something which persists throughout the change.

So, while Kant is entitled to claim that through the fact of reason, we have (non-theoretical) access to some noumenal reality, this is only under the category of cause and effect, and there are no grounds for talking of a noumenal self understood as a substance.[8]

The issue of whether Kant himself saw things this way is of course more unclear. In this passage, Kant indeed refers to the 'subject as a thing-in-itself'. Additionally, the postulates of practical reason directly point at a substantial notion, namely the traditional concept of soul. Thus Kant claims "[t]his infinite progress [to perfect fitness to the moral law] is possible, however, only under the presupposition of an infinitely enduring existence and personality of the same rational being" (CPrR: 122). Looking at Kant's very brief justification for this claim, we note

[8] There is more to say on this issue, because Kant refers elsewhere to our having grounds for considering ourselves as members of an intelligible world. This issue is examined below.

that he validly moves from the claim that "[t]he achievement of the highest good in the world is the necessary object of a will determinable by moral law" (CPrR: 121–2) to the claim that "it is necessary to assume such a practical progress as the real object of our will" (CPrR: 122). But in all this proof, no grounds are adduced for viewing this will as a faculty of some noumenal substance, so the introduction of the idea of the soul could only be justified on an understanding of the soul as defined by its willing. And the will, for Kant, is the faculty through which we are causally effective.

What is arguably lurking here in Kant's text is a default assumption that typifies much of Western metaphysics as Heidegger (1962: 72–3) would point out, namely that when thinking about being, the category of substance is primordial. Without pursuing this question, all we need to note here is that considerations about the will of 'rational beings' provide no warrant for referring to them as noumenal substances.

Noumenal transcendental freedom?

Second, the notion that the noumenal self is spontaneous presents us with the challenge of making sense of the transcendental freedom of a noumenon, an issue that a number of authors have commented upon. There are two possibilities here:

(i) the noumenal self contains characteristics accounting for both our adherence to and our infringement of the moral law; and
(ii) the noumenal self only acts morally.

Possibility (i) contradicts Kant's repeated assertion that there is no 'law of evil', i.e. that an agent does not decide against what the moral law commands on the grounds that he adopts a principle of evil (see e.g. AAVI: 43). On such an understanding of evil, this cannot therefore be a feature of the intelligible character as a subject's noumenal causality. So it would appear that the intelligible character is morally good, as Kant's writings on practical reason also suggest (e.g. GMM: 455; CPrR: 43, 47).[9] This means adopting option (ii), but this leaves it completely obscure how it might be possible to act other than according to the moral law which is a law of the intelligible world. This incomprehensibility is not just that which Kant flags, namely the impossibility of making sense of how an agent would decide to act against the moral law

[9] The Third Antinomy does not, however, suggest that this is the case. We shall return to this difference further.

(e.g. AAVI: 43). That we cannot make sense of *why* our power of choice would opt for a maxim forbidden by the moral law because we cannot see in what sense this could be deemed rational is one problem, but a problem which might seem "confined" to Kant's ethics. But the further problem of its not being understandable *how* such action is possible at all, given it is the result of a morally good intelligible character's power of choice, looms large and threatens the cogency of Kant's notion of transcendental freedom.

Indeed, this problem can also be expressed in terms of freedom. If, as Kant says in the second *Critique*, "supersensuous nature, so far as we can form a concept of it, is nothing else than nature under the autonomy of pure practical reason" (CPrR: 43), freedom is thereby viewed as the law of this nature, which seems to make it incomprehensible how one could freely fail to act morally. Initially, Kant was criticized by some of his most fervent immediate followers for appearing to make all freedom of choice disappear (e.g. Reinhold 1790–2). But soon, those who took it upon themselves to "complete" the Kantian system, and today, many interpreters of Kant, have adopted the line that this is, respectively, the most fruitful, or the most accurate interpretation of Kant's texts (Kosch 2006):

(i) The focus upon freedom as autonomy, i.e. as self-determination, marked the development of German Idealism in its early stages; thus Schelling's *System of Transcendental Idealism* (1800) argues for a notion of freedom as self-determination through which individuals, insofar as they participate in the absolute, bring about a world governed by intelligible principles.
(ii) Contemporary interpreters of Kant's ethics such as Korsgaard (1996: 159–87) and Wood (1999: 172ff) take the view that there is a failure of practical rationality involved in choosing to act immorally, i.e. that such action is never a proper instantiation of freedom.

That such contemporary interpretations should be attractive on merely ethical grounds is arguable, and to some extent, they may seem to be called for by Kant's claim that one can never choose evil *per se*. There are passages in Kant's texts suggesting that he took this view (e.g. in the *Metaphysics of Morals*, AAVI: 226), but as Kant also always reminds us of the imputability of evil action, this cannot be his considered opinion. And, indeed, one must not ignore the fact that Kant's notion of transcendental freedom, as it is first set out in its theoretical context, emphasizes the freedom to have done otherwise, which is indeed what is experienced through our understanding of oughts (A547/B575). With

this notion of freedom, the moral law is not characterized as *descriptive* of the causality of the intelligible domain, but rather as *prescriptive* for rational agents subject to inclinations (Kosch 2000: 46–50). Such a notion of freedom can precisely address the issue of how it is that we do not always act morally.

The focus of German Idealism upon freedom as self-determination was eventually to be criticized by Schelling himself, and the demise of German Idealism can be seen as related to the problem Schelling flagged. As Schelling pointed out in the *Freiheitsschrift* (hereafter FHS, 1809), the very notion of evil seems to disintegrate if failing to act morally is to be understood exclusively in negative terms, i.e. in terms of inclinations coming to dominate our capacity for freedom (FHS: 371). And here again, Kant himself provides clear evidence that a notion of evil, even radical evil, has a role to play in human agency. The *Religion within the Limits of Reason Alone* text introduces three degrees of evil. First, there is frailty, which involves the agent succumbing to strong inclinations. Second, impure motives arise when duty is not the motivation for adopting some maxim compatible with the requirements of the moral law. Third, depravity involves subordinating the moral law to non-moral motives (AAVI: 29–30).

Wille and Wilkür

Bringing the two points together, the *Religion* is a text in which Kant provides an account of freedom that includes both the focus upon self-determination according to moral laws, and the freedom to do otherwise. Kant has ways of accounting for the volitional aspects of such evil action which involve the introduction of the distinction between *Wille* and *Willkür* (AAVI: 226). The former is the legislative capacity of the will (also called '*Wille*'), while the latter is the power of choice. Additionally, with the Incorporation Thesis in the *Religion* (AAVI: 23–4), as Allison (1990: 40) calls it, Kant clearly states that an incentive can only determine action if it has been incorporated into the agent's maxim. This means that, even in the case where an agent 'gives in' to some inclination which is contrary to the requirements of the moral law, a free choice has been made, namely that of letting the inclination in question determine one's agency. With this notion of incorporation, a text such as MM: 226 no longer appears to support the view that immoral action is not free: the issue is rather one of grasping how the agent can opt for acting against the law as there is no rational principle which can explain such a choice.

But does this resolve the metaphysical conundrum of how a good intelligible character could co-exist with immoral action? Apparently

not, since the *Religion* text does not explicitly deal with metaphysical issues arising from the notion of freedom. What is needed is some metaphysical underpinning for the *Wille/Willkür* distinction.

Nor are this notion of *Willkür* and the metaphysical question it raises only a feature of the *Religion* text. In the discussion of the Third Antinomy, Kant distinguishes practical from transcendental freedom: practical freedom (A534/B562) involves a power of choice which is independent of determination by natural causes (A533–4/B561–2). This power of choice is the power to do or not do as I ought. For Kant, it is a practical notion, but that leaves untouched the question of the nature of that which is endowed with practical freedom: it cannot be a noumenal self, endowed with an intelligible character, for its actions must be able to diverge from the latter's laws of freedom. It cannot be an empirical self because the latter is embedded in the causality of nature, and so any freedom which it could be described as having would be merely psychological. So we find ourselves with a similar question to that encountered with the notion of *Willkür*.

Another notion of freedom Kant sometimes adduces is that of negative (transcendental) freedom (AAVI: 226) which is, in effect, just this power not to have done what one did. This characterization seems to belittle its importance and present it as merely a negative aspect of transcendental freedom. But while this power of choice is of course reflected in the empirical character, it is a transcendental characteristic of our freedom (positive freedom requires negative freedom to be efficient) that is not a feature of the intelligible character. This therefore leads to the question of what it is that is endowed with such negative freedom or a *Willkür*.[10]

3 The metaphysics of Wille and Willkür and Sartre

As we have just seen,

(i) it may be thought that, insofar as *Wille*, in the narrow sense, is practical reason, i.e. the will under the moral law, and therefore a noumenal feature, *Willkür* by contrast is the will of our empirical selves; but *Willkür* is that faculty which is transcendentally free

[10] One might want to answer that these metaphysical questions are just a reflection of the lack of a clear theory of the self in Kant, for it is after all the self which is the bearer of these properties. I think that while these questions do indeed arise because there is no such theory, there are also good grounds for Kant's not having such a clear theory (Onof 2010). Consequently, it is not a unifying theory of the self that I shall be proposing here to address these metaphysical questions.

in the negative sense, i.e. in the sense that it is that faculty which enables us to do otherwise than what is prescribed by any form of determination, while *Wille* is the faculty of acting morally, i.e. positive freedom (so that both together define autonomous action: autonomy is a property of *Wille* in the broad sense). So *Willkür* cannot be the property of an entity that is caught up in the deterministic world of empirical causality.

(ii) Equally, as a feature of the will which is transcendentally free, *Willkür* may be thought of as a feature of the noumenal aspect of the self. But then, as we saw, the question arises as to how there could be a noumenal choice to act against the moral law. This is an issue we examined above: since the law of the noumenal is the moral law, no possibility of noumenal infringement of this law exists.

Spontaneous consciousness and nothingness

Sartre's 'realist' ontology of the In-Itself is, of course, fundamentally different from Kant's transcendentally idealistic metaphysics, largely as a consequence of Sartre's ignoring the kind of epistemological questions which gave rise to Kant's claim of the transcendental ideality of space and time. Sartre's For-Itself would find no plausible home either in the world of appearances or in Kant's domain of the in-itself insofar as it is a way of being defined by a lack of being (and a lack of identity), all of which have no place in Kant's metaphysics. This feature could, however, be of use to interpreting the metaphysical status of *Willkür*.

Indeed, the key feature of Sartre's ontology which enables him to defend his strong notion of freedom is his locating the For-Itself outside the fullness of being of the In-Itself. It is only because the For-Itself stands outside being In-Itself, and is in fact separated from it by *nothingness*, that the For-Itself is free in an unconditional and absolute sense.

The suggestion from Sartre's ontology is therefore to consider our spontaneity as located outside the domain of being which these two worlds define. Given this parallel, we might therefore consider whether it is plausible in terms of Kant's system, and indeed desirable, to consider the negative transcendental freedom of the spontaneous agent as a feature that is external to being. In what follows, I shall refer to 'spontaneous consciousness' (hereafter SC) as that entity endowed with *Willkür* that we are, i.e. as that which has the property of negative transcendental freedom.

With this proposal, one should immediately be reminded of the duality of Kantian distinctions appearances/things-on-themselves and empirical/intelligible, whereby in the latter distinction, the intelligible is what is considered by pure reason independently of any intuition, and

does not claim to refer to anything, while the first distinction certainly refers to things as they are in themselves as beings for all but the methodological two-aspect theory propounded by Allison and Bird. And this is where considering our SC as (at least partly) outside the domain of being emerges as a possible interpretation. That is, we can indeed have an intelligible conception of this entity that we are, and on this conception, it is an entity that is not located in either the realm of appearances or the realm of the In-Itself.

While such a location of the SC outside these domains of being is motivated by the two issues we have examined, namely the problem of understanding the possibility of acting against the moral law and the fact that there are no grounds for viewing any 'substance' self as belonging to the intelligible domain, there are further important grounds for it.

This status of the entity endowed with *Willkür* is in fact required if we want to make sense of how we could belong to two different worlds or domains, while being a self, and therefore being in some sense unified. Kant was much criticized by his followers, in particular Schiller and Hegel, for having a view of the self according to which the Kantian self is torn between its duty, i.e. the law that describes its noumenal reality, and the pull of inclinations in the phenomenal world (Onof 2011a). It is not clear that Kant has an answer to this question, and, indeed, when this question is taken as an ethical one, it provides a powerful source of objections to Kant's approach, as we see for instance from Bernard Williams's reformulation of a broadly Hegelian type of critique (Onof 2011a). However, if we now view the kind of entity that we are insofar as we are SC, as located neither within the world of phenomena or noumena, this entity can indeed have a unity in time that comes from the kind of decisions that it makes over a lifetime.

Gesinnung

And Kant has an account of the unity of these decisions in the form of his notion of *Gesinnung*. This fundamental disposition of our moral character is chosen by the agent. This notion of choice is not an easy one to make sense of, and it has been suggested in particular that there is an instability between its playing a role of Aristotelian ἕξις (moral disposition) and something like Sartre's 'projet fondamental' (O'Connor 1985: 293): that is, it both reflects the continuity of a person's moral character and is freely chosen. While there may appear to be a tension between these two understandings of this notion of moral disposition, they can be reconciled if the *Gesinnung* is understood as a fundamental maxim underpinning our moral choices (Allison 1990: 142; Kosch 2006: 61).

This is supported by Kant's characterization of it as the first subjective ground of the adoption of maxims (AAVI: 25). What is noteworthy here is the notion of a subjective ground: this choice of disposition cannot be a choice made at the noumenal level because it would not be subjective, and it cannot be a choice in the phenomenal world because there is no freedom in the causally determined empirical world. This is the choice of an entity that is not located in either of these worlds, i.e. is not that of any being, but is, through its choice, directly related to a phenomenal and a noumenal being. Moreover, by thus being related to both, this choice arguably defines a notion of unity of the self.[11]

Additionally, this understanding of *Gesinnung* reflects the metaphysical nature of the agent's SC: it is neither in time (phenomenal) nor timeless (noumenal) but related to both, so that one can understand the agent's spontaneity in terms of a single fundamental choice (timeless) that underpins all her actions, or in terms of discrete choices made in time, depending on whether we focus upon the relation of this spontaneity to the noumenal or the phenomenal worlds.[12]

4 Spontaneity and the intelligible and empirical characters

To fill in this notion of spontaneity by reference to the notion of *Gesinnung* leaves unclear in what these relations to the noumenal and phenomenal worlds established by our free-standing practical

[11] Another interpretative option that avoids such metaphysical issues is, of course, to adopt a methodological two-aspect theory of the relation between the in-itself and appearances (Allison, Bird), and indeed the defence of this interpretative approach draws upon the metaphysical complications that such Kantian notions give rise to, to justify their thesis that Kant's claims cannot be understood as metaphysical (e.g. Allison 1990: 140).

[12] I don't take this analysis to remove the difficulties involved in understanding the notion of *Gesinnung*. Baldwin (1980) tries to provide an interpretation of the relation of both Kant's notion of *Gesinnung* and Sartre's Projet Fondamental to individual choices, as a process analogous to how a painting is produced as the result of a number of brushstrokes. Baiasu (2011) rejects this analogy, chiefly on the grounds that it cannot account for how each action makes sense against the background of *Gesinnung*/Projet Fondamental (G/PF), since the latter is essentially indeterminate at any particular point in time. Baiasu correctly identifies issues with an interpretation of the choice of G/PF as the result of deliberative choice, but this does not imply that the painting analogy is inadequate (Onof 2013a). The metaphysical account proposed here suggests such a picture, with the individual choices made by our practical spontaneity contributing to a further determination of the empirical character partly through the intelligible character.

spontaneity actually consist.[13] To shed some light on this, we need to consider what Kant says of the relation between the intelligible and the empirical characters. Kant talks both of the empirical character as the appearance of the intelligible character, and of it as its causal effect (Allison 1990: 32). In fact, both descriptions are quite compatible if we view appearances as the result of affection by things in the domain of the in-itself.

However, to accept this does not tell us exactly how our spontaneity might be involved. What Kant does tell us about our spontaneity is clearly expressed in the Incorporation Thesis (Allison 1990: 38–9).[14] This thesis can be understood as a model for the way our practical spontaneity operates, and the analogy with our theoretical spontaneity is well brought out by Allison's analysis: the agent's practical spontaneity consists in her choice to incorporate an incentive into her maxim of action (for Kant, the choice is essentially inclination or duty) by *taking it* as ground for action, in the same way as an object *x* is *taken as* F when we judge '*x* is F'. As above, one can understand this act as involved in the choice of one's fundamental maxim, and in the choice of each maxim of action in the temporal order of appearances.

In terms of causality, what this incorporation amounts to is *prima facie* a choice between letting one's action be determined by the law of freedom or letting natural causality in the guise of some inclination take its course (but this is still a free choice we make). This explains Kant's many statements in which he suggests that infringing the moral law is not an act of freedom: what is meant is that in choosing inclination over duty, one is giving precedence to the causality of nature over the intelligible causality (that of the moral law). In thus exercising my spontaneity, I can bring some aspects of what is thereby identified as my intelligible character to appear in the phenomenal world in the form of subjective maxims characterizing my empirical character. That is, my spontaneity as a transcendentally free agent consists in enabling some aspect of that which is thereby identified as *my* intelligible character to appear in the phenomenal world.[15]

---

[13] I focus upon the relations of practical spontaneity. In its theoretical employment, spontaneity relates to being in different ways. We shall come back to this point at the end of the chapter.

[14] I shall not adopt this as a model for the intelligible character, but rather for how the agent's spontaneity brings the intelligible character to appear in the form of a further determination of the agent's empirical character.

[15] The personal aspect of this choice can be given further substance if one considers that each agent will have to make choices as to how he ranks duties, choices which are required by the circumstances encountered in the phenomenal world, but I shall not defend this interpretation here (see Onof 1998).

Here, we should note that if we look more closely at the role of incorporation, the situation turns out to be more complicated. One can act from an inclination which is not the strongest one. Such action could be immoral or permissible; but in either case, it will require some non-natural, therefore intelligible, causality. The intelligible character can be understood as playing a secondary role in this type of action; to avoid complicating the account, I shall ignore this case in the rest of the chapter.

The role of the agent's spontaneity is therefore to be able to bring about the relation of appearance between the intelligible character and the natural order. Unlike other things-in-themselves which appear in the natural world through the relation of affection (Onof 2011), what is at stake here is the appearing of a certain form of causality (according to the moral law) in a world governed (self-sufficiently) by the natural laws of mechanistic causality. It should therefore come as no surprise that it is only through some external 'intervention', namely that of the agent's spontaneity, that this should come about.

By drawing upon Sartrean ontology, I have proposed that our practical spontaneity (SC) be understood as lying beyond being (in-itself or appearances)[16] but related to it in ways defined by the agent's disposition (*Gesinnung*). This relation determines how and when this practical spontaneity as *Willkür*, in its negative transcendental freedom, brings the causality of the intelligible character (which follows the moral law of the *Wille*) to bear upon the world of appearances in the form of further determinations of an empirical character. This amounts to a realization of positive transcendental freedom in the world of appearances.

5 A threat to the possibility of freedom in Kant's system

Statement of the problem

In the CPrR, Kant revisits his solution to the free will/determinism problem from the first *Critique* (CPrR: 97–100). As noted above, the tone is different now, so that the account of the compatibility of a deterministic phenomenal world with transcendental freedom in terms of a duality of empirical and intelligible characters is no longer presented as

[16] This dichotomy does not suggest a two-world view of being, but is meant to be non-committal as to whether a two-world or another view of the relation in-itself/appearances is taken. I have argued for a two-perspective view elsewhere (Onof 2011).

theoretical speculation, but an account for which practical grounds are given by the fact of reason (CPrR: 99).

But there is another important difference in what follows this passage in CPrR. Namely, Kant raises a new type of difficulty that was not discussed in CPR, "a difficulty which threatens freedom with its complete downfall" (CPrR: 100). The difficulty Kant identifies is, he tells us at the outset, that it would be greater were it the case that "the existence in time and space is held to be existence of finite things in themselves" (CPrR: 100), thereby suggesting that Transcendental Idealism will provide a solution.[17]

The problem Kant identifies is that "as soon as it is assumed that God as the Universal Primordial Being is the cause also of the existence of substance ... one must also grant that the actions of man have their determining ground in something completely beyond his own power ... upon which his existence and the entire determination of his causality absolutely depend" (CPrR: 100–1).

Prima facie, the difficulty seems to be that, even if one can appeal to an intelligible character that is transcendentally free, the fact that this character with all its causality is in fact entirely created by God leaves it unclear to what extent it is *my* freedom that is responsible for my actions: the "determining ground" of my actions would seem to have been chosen for me by God. I shall call this Problem A.

It is not clear how Kant can address this difficulty if he takes the "determining ground" of my actions to lie in my intelligible character, and therefore to lie in what God creates as my intelligible character. In fact, this is not, however, how Kant understands this difficulty: his focus is entirely upon the "actions of man" (in the phenomenal world, CPrR: 101) and how these would be determined by God. This does indeed define a problem (hereafter Problem B), but I shall argue, solving this does not resolve the other one: these are two distinct, if related, issues.

Problem B, which Kant tackles, can be understood by recalling that the solution Kant presented in the Third Antinomy relied upon the fact that the causal account of my action in terms of determining causes

[17] One might wonder why this problem was not addressed in the second edition of the CPR, unless it is a problem Kant stumbled upon only in 1788. More plausibly, one should note that the problem arises only on the assumption of God as the universal primordial being. The regulative status of this assumption for theoretical reason is however weaker than its regulative status for practical reason: in the CPrR, it will become a Postulate of Pure Practical Reason which one is committed to through one's moral agency (CPrR: 124–32).

in the world of sense does not exclude a complement of determinacy: as in all causal regresses, one can seek grounds for the causality of any cause, and one can therefore seek further grounds for any antecedent cause being causally effective in determining my action, i.e. through determinations of my empirical character, and this is where an appeal to my intelligible character makes sense. What Kant is now pointing out is that it would seem that all such causal regresses have to refer back to the creation of all substance by God, thus leaving no room for any complement of determinacy from my own transcendental freedom. Thus Kant suggests, "self-consciousness would indeed make him [, i.e. the agent,] a thinking automaton, but the consciousness of his spontaneity, if this is held to be freedom, would be a mere illusion" (CPrR: 101).

Kant's solution to this problem is to point out that "the creation of beings is a creation of things in themselves … creation concerns their intelligible but not their sensible existence, and therefore creation cannot be regarded as the determining ground of appearances" (CPrR: 102). This solution is certainly not obvious and Kant has little more to say about it, apart from drawing attention to the (much worse) problems that would arise were things in time to be taken to be things-in-themselves.[18] I think Kant's solution is intriguing and that it can also provide a powerful way of dealing with the theological problem of evil, but I shall leave this topic aside for the purposes of this chapter.

What is not clear is why Kant thinks that it is harmless that my intelligible character should be entirely determined through creation, i.e. Problem A which I flagged above. Where would freedom lie if the intelligible character were determined by divine creation? Kant is not trying to avoid the problem as he clearly states, in addressing the other problem, that "creation … concerns their intelligible … existence" (CPrR: 102).

In effect, the issue we are faced with could be seen as closely related to the worry discussed earlier, that, if the intelligible character can only cause moral action, which is indeed what must presumably be the case if it is created by God, freedom of choice is not located therein.[19] What this passage in the CPrR confirms is that it must be Kant's considered view that the intelligible character is just a causality according to the moral law, and this passage adds that this character must be created by

[18] Kant shows that because God could not create anything without time and space already being available, his causation of the existence of these things would be conditioned (CPrR: 101).

[19] It is, in fact, a bigger worry, for here it would appear that there is not even any scope for individuated choices of implementation of broad duties for instance (see footnote 15).

God, and these claims must cohere with its being the case that he did not see this as threatening our freedom of choice. But how can these views be reconciled?

On the interpretation I propose in this chapter, our spontaneity is not a feature of any being, but SC is located outside being, and its relation to an intelligible character is one through which it further determines the empirical character, which leaves conceptual space for freedom and addresses Problem A.[20] With this solution to Problem A, freedom might now appear to be reduced to the freedom *not to do* what is moral. This hardly seems satisfactory given Kant's understanding of freedom; additionally, without the interpretation I propose, Kant would apparently not have any space for freedom at all. In fact, looking at this more carefully, the Kantian response to Problem A must involve differentiating intelligible causality *in general*, which is created by God, and *which* intelligible character I choose for myself. That is, my spontaneity's drawing upon intelligible causality to further determine the empirical character must be an ongoing process through which the intelligible character is *made mine*: even assuming that God is the ground of all intelligible causality, the making mine of an intelligible character is a further issue which is entirely up to me, and a feature of my freedom. Consequently, my interpretation needs to specify that my relation to an intelligible character is a *making mine* of a certain intelligible causality (which can only therefore be described as *my* intelligible character insofar as I make it mine).

Can we draw from Sartre's ontology to address this Kantian problem?

What is now required is an understanding of the nature of this SC that is located neither in the phenomenal nor the noumenal world.[21] The fairly

[20] One might ask about the genesis of SC. I think this can be viewed as compatible with Kant's practical belief in a God that is responsible for all creation, because what is created here is located outside being, and as we shall see below, is rather to be viewed as structure of possibilities. This would, however, require the nature of God in relation to being to be examined more closely.

[21] More precisely, in the Third Antinomy, the issue is of showing that this is not impossible, while in the CPrR, it is rather an issue of providing a model showing how it is possible. Arguably, the first task is already achieved by Kant even without resolving the ontological problem that the notion of *Willkür* gives rise to. But, for the investigation of the CPrR, something needs to be said to make sense of Kant's resolution of the additional problem that he identifies for the possibility of freedom.

obvious inference, that it is a nothingness, brings us back to looking at how Sartre construes this notion.

As we saw above, Sartre's notion of the For-Itself is defined as a disturbance, "something which happens to the In-Itself" (BN: 216) that has no place in the fullness of being which is the In-Itself. Without pausing to consider how congenial such an ontology might be to the overall picture of Kant's system, let us note that there is a key feature of this For-Itself which appears relevant to the Kantian picture.

This is that reality does not simply consist, for Sartre, of the In-Itself in its self-sufficient existence and the For-Itself which is as a disturbance in the fabric of being, something which happens to the In-Itself, indeed, the only thing that happens to the In-Itself, or, as Sartre describes it, "the only possible adventure of the In-Itself" (BN: 216). There are ontological relations, and consciousness is always a relation to an object of consciousness: "All consciousness, as Husserl has shown, is consciousness of something. This means that there is no consciousness which is not a positing of a transcendent object" (BN: xxvii). Such ontological relations provide the foundation for epistemological relations for Sartre, which partly explains what might seem to be a neglect of epistemology: Sartre accepts that "the being of the For-Itself is knowledge of being", but this follows from 'the fact that in its being it is a relation to being' (BN: 216).

What are these ontological relations which support knowledge? Sartre gives pride of place to the concept of negation and implicitly refers to Kant when he claims that "it is not true that negation is only a quality of judgment" (BN: 7). Rather, negation presupposes an attitude of questioning which "is a relation of being" (BN: 7). In this questioning relation, what is expected is not a judgement, but a "disclosure of being on the basis of which we can make a judgment" (BN: 7). And expecting such a disclosure also means being prepared for "a disclosure of non-being" (BN: 7). Sartre gives several examples of cognitive enquiries to show how negation is involved in the questioning attitude, and he does not remain in the purely theoretical domain, but also considers our practical undertakings (BN: 7). The attitude of questioning thus reveals the negating power of consciousness so that negation can properly be described as a transcendental condition of all forms of cognition (Gardner 2009: 62).

To convince the reader that such a condition is indeed ontological, Sartre shows in what way negation defines real modes of being, such as that of being absent. Famously, he considers that Pierre's absence from the café is manifested in experiencing the café without Pierre: Sartre talks of an "intuitive apprehension" (BN: 10). The Kantian language is meant

to show in what way Sartre's theory is designed to supersede Kant's transcendental conditions by "providing a unified theory of understanding and sensibility: negation becomes a single transcendental condition of conceptuality (of judgement and concept application) and of perception (negation provides a form of intuition not unlike but even more basic than those of space and time)" (BN: 65).

This is of course a very un-Kantian type of transcendental condition, chiefly because it is an ontological one, while Transcendental Idealism is primarily concerned with conditions for knowledge of objects. Nevertheless, it is an open question whether Kant's doctrine does not inevitably lead to ontological issues which he himself touched upon, e.g. the existence of things-in-themselves (Onof 2011) or which others took up after him (e.g. Schelling's *System of Transcendental Idealism*). In particular, for our purposes, there are at least two transcendental conditions which, I claim, give rise to further ontological questions.

Focussing on the spontaneity of the understanding for Kant, the highest condition of all knowledge is the condition that "[t]he I think must be able to accompany all my representations" (B131). This is a formal requirement, but it does, however, refer to a spontaneity that is able to produce the representation 'I think'. That Kant should not have discussed this ontological issue is germane to his purpose of identifying the transcendental conditions of knowledge. There are, however, further conditions that must remain inaccessible to our cognition, as Kant readily recognizes in the case of Schematism, for instance (A141/B180–1). It would seem that the nature of that which supports the required spontaneity of thought is similarly beyond our ken.

It does, however, become relevant when the issue of the nature of our practical spontaneity arises, as the discussion above has suggested. And, without endorsing the identification of the spontaneity of the 'I think' with our transcendental freedom (see Prauss 1983), it is clear that the same spontaneity of thought is involved in both, as Allison's interpretation (1990: 37–9) of the Incorporation Thesis shows: practical spontaneity involves taking an incentive as ground for one's action in the same way as the understanding takes *x* as *F* when it determines an object under concept *F*. And this is not only an analogy: practical spontaneity requires theoretical spontaneity, i.e. the judgement that an incentive is a ground for action.

If this suggests that Sartre's ontology of the For-Itself could be useful to supplement Kant's transcendental structure of practical spontaneity with an ontological basis that would at least not be self-contradictory (and thus fulfil the requirements of the Third Antinomy), there is a

more general point that emerges from his doctrine of consciousness as nothingness, i.e. as that which is the source of all negation, which suggests some degree of compatibility between Sartre's doctrine of intentionality and Kant's understanding of objectivity. This is the conclusion Sartre reaches when he revisits his interpretation of consciousness in the light of his extensive examination of the structure of the For-Itself. He now clearly states that the original negating act of consciousness is that through which consciousness is defined, i.e. it is defined in relation to the being which it negates: "The For-Itself is a being such that in its being, its being is in question in so far as this being is essentially a certain way of not being a being which it posits simultaneously as other than itself" (BN: 174).

6 Spontaneity and negation

Spontaneity as a nothingness

I shall briefly examine grounds for viewing an understanding of the ontology of our spontaneity characterized by a negation of the object it is intentionally directed to, as compatible with Kant's theoretical and practical philosophy. These are, first, the notion of transcendental object in its relation to the transcendental subject in the A-Deduction, and second, the role of the 'ought' in moral deliberation.

With the concept of the transcendental object introduced in the A-Deduction, Kant wants to explain how "our empirical concepts in general can provide relation to an object, i.e. objective reality" (A109). This concept plays the role of a placeholder, i.e. "something in general = X" which is a condition for being able to speak "of an object corresponding to and also therefore also distinct from the cognition" (A104). This is a key component of Kant's account of objective knowledge, and it crucially involves a negation insofar as it is defined through *differentiation* from our representations: something in general = $\times \neq$ our representations. Just as the unity of the synthesis of the manifold is not a determination under the category of unity, the differentiation of the transcendental object from our representations is not a determination under the category of negation: rather, in both cases, we have a condition for the possibility of any objective determination, i.e. any determination in terms of quantity, quality and relation, which is part of the structure of transcendental subjectivity.

Turning now from the theoretical to the practical dimension of our subjectivity, as we saw above, the Third Antinomy contains an analysis

of whether we can make sense of the notion of transcendental freedom, i.e. the causality of reason. Kant considers the fact that we guide our action using imperatives, in which the key notion is that of the 'ought'. As Kant puts it, in nature "it is impossible that something in it ought to be other than what ... it in fact is; indeed, the ought, if one has merely the course of nature before one's eyes, has no significance whatever" (A547/B575). Kant adduces the existence of practical oughts to support his claim that it is possible to represent that reason has causality. This causality is manifested in the fact that "with complete spontaneity it makes its own order according to ideas ... according to which it even declares actions to be necessary that yet have not occurred and perhaps will not occur" (A548/B576). In formulating such an order reason implicitly negates the actual order of things (past, present or future) in nature in which the 'ought' has no place.

We therefore have two moments, involving an essential role for negation, characterizing, respectively, the spontaneity of the understanding in its theoretical relation to the world and the spontaneity of reason in its practical relation to it. This enables us to draw a parallel with Sartre's assigning a key role to negation in defining the being of consciousness, and to make sense of how our spontaneity could be said to be a nothingness.

There are, however, also grounds for resisting such an interpretation. First, much as transcendental idealist interpretations can provide a cogent interpretation of much of Sartre's metaphysics in BN (Gardner 2009: 73–80), Sartre rejects idealism (BN: xxxviii). The key point is that Sartre does not accept that the objects of our conscious experience are constituted by our subjectivity. Additionally, the negating which seemingly operates as a Kantian transcendental condition defines a world that, while it is intersubjectively accessible, is perspectival for a particular subjectivity: the absence of Pierre from the café is not part of John's world insofar as John is not acquainted with Pierre. So the role of negation in Kant's understanding of the relation to the object is clearly distinct from Sartre's. Second, on a more general level, the Sartrean emphasis upon negation and nothingness is of a piece with an overall picture of the human condition as characterized by the hopeless attempt of the For-Itself to make up for what it lacks by becoming an (impossible) In-Itself-For-Itself, i.e. to become God (BN: 566). This is clearly in tension with the overall thrust of Kant's system, and in particular with his regulative idea of progress towards the Highest Good. Without further examining this issue, I will suggest grounds for not following Sartre in his exclusive emphasis upon negation.

Outline of an alternative proposal to account for spontaneity

Looking a bit more closely at Sartre's choice of examples to illustrate the role of negation, we might want to consider a somewhat different approach to understanding the nature of the spontaneity of consciousness. The famous case of Pierre who is absent from the café is an example of determination which, in Kantian terms, *happens* to be under the category of negation (Pierre is 'not' in the café). But how essential is this negative form? Sartre is certainly right to indicate that referring to the category of negation does not provide a sufficient explanation of the nature of the negation here, because negating Pierre's presence is not the same as negating the presence of Wellington in the café (BN: 10). But what is missing in the purely categorial notion of negation is that there are, or were, times at which Pierre would have been *expected* to be in the café. Pierre's presence was therefore experienced as belonging to an understanding of the café, which understanding involves defining relations between what there is and what is expected.

Now such expectations express the existence of an *ought*, but a cognitive as opposed to a practical one. In Pierre's case, what is implied by saying that he is not in the café is that there are circumstances under which he would probably be found in the café. If these circumstances are known to be realized, then the statement that Pierre is absent expresses surprise and this can be rendered as 'he ought to be there'. Sartre uses other examples, e.g. of a car breaking down or a watch stopping (BN: 7), to express similar types of surprise. In other cases, where the circumstances are not known to be realized, there is no surprise, but these only make sense by reference to the first type of circumstances and the associated notion of what is most probable.

So, although the judgements in question come under the category of negation, their essential feature is, as Sartre's text itself sometimes suggests, that they are about expectations, and therefore about cognitive oughts, i.e. about normativity. As with the practical oughts which, in the Third Antinomy, Kant presents as characteristics of our practical freedom, cognitive oughts only make sense on the assumption that it is possible that things should be different from what they are. So if there can be any normativity (cognitive, practical or of some other type), a grasp of *what is possible* is required. The proposal would therefore be to understand the nature of our consciousness[22] (for Sartre) or our spontaneity (for Kant) as

[22] In the Sartrean account, I would argue that it is self-consciousness that is thus to be interpreted as possibility, but this point is academic since all consciousness involves self-consciousness for Sartre.

possibility rather than *actuality*, so that the origin of possibility is thereby accounted for.[23] That is, rather than understand possibility as arising with the For-Itself (BN: 99), and therefore as a consequence of its being nihilation, as Sartre does, I propose taking it to be the primary characteristic of consciousness. A particular consciousness (SC) is a particular structure in the 'domain of the possible'. This structure is normative, and defines the possibilities which are closer or more distant for me; this organization of the possible defines how the world appears to me and what the motivations for my actions are.

In terms of the appropriateness of this shift from negation and nothingness to the modality of possibility to explain the ontological status of Sartrean consciousness, I will just mention two considerations, since the purpose of this chapter is not a critique of Sartre's theory:

(i) Sartre himself criticizes Heidegger for the use of positive rather than negative terminology to describe Dasein's transcendence (BN: 18). This criticism can, as shown above, be turned back on Sartre's excessive emphasis upon negative terms.
(ii) Sartre's account of possibility rightly involves a rejection of any realist account of it in terms of being In-Itself. Sartre then defines it in terms of the lack which characterizes every For-Itself (BN: 102) and thus in terms of nihilation. This step follows from the fact that possibility is understood as an attribute of consciousness, i.e. an "*a priori* feature of human subjectivity" (Gardner 2009: 106). This leaves it open that there is a more fundamental role of possibility as definitive of the ontologico-modal status of the spontaneity of consciousness. While this may seem to deny an obvious actuality of such conscious experience, this actuality is accounted for by the fact that consciousness is always of something (BN: xxvii) which can be interpreted as consciousness as possibility always being directed to being and thereby defining what is actual.

Such a "*modal realist*"[24] proposal will nevertheless seem at least as strange as Sartre's identification of consciousness with nothingness (BN: 23), and inappropriate to fill in the Kantian metaphysical account. Here,

[23] Sartre understands possibility in terms of the nihilations of consciousness; I propose to invert the explanatory order.

[24] I use the word 'realist' insofar as the domain of the possible is here understood as part of reality, though it is outside the domain of being (either the in-itself or the world of appearances).

I think it is important to add that the possible is nothing in isolation: it is always a possibility of being, i.e. it is directed to being, which thereby accounts for the intentionality of consciousness. Consciousness's directedness to being thereby defines what is *actual* in the world of appearances, in ways which we shall examine below. This relation inverts that characteristic of David Lewis's (1986) modal realism, in which possible worlds are alternative worlds to the actual one, whereby the possible is defined in terms of the notion of actuality.

With this alternative proposal to Sartre's identification of consciousness with nothingness, we nevertheless import into an interpretation of the metaphysical status of spontaneity for Kant something which is beyond the framework of Transcendental Idealism. As the purpose here is just to outline the proposal, I shall focus upon first showing that it is not incompatible with Kant's understanding of practical normativity, and that it can be seen to receive support from his theory of theoretical normativity.

Prima facie, understanding the spontaneity of an individual consciousness in terms of a structuring carried out in the domain of what is possible in relation to being is foreign to Kant's conceptions of practical freedom. However, in Sartrean terms, this could be viewed as defining the agent's fundamental project insofar as such a project just amounts to an ordering of the possibles involving human agency around the project's conception of my life.[25] And, therefore, using the close connection between Sartre's fundamental project and Kant's *Gesinnung* (Baiasu 2011: 73–5), this ordering of possibilities involving human agency can be viewed as structuring our practical life, and could be translated in terms of those maxims of action characterizing a particular agent.[26]

[25] The possibility which the For-Itself is defines the 'Circuit of selfness' (BN: 102), which is one of the two core aspects of Sartre's notion of self (Onof 2013). Since the fundamental project is defined in terms of the For-Itself seeking to ground itself, this must amount to identifying what is possible in relation to this goal, and which possibilities are closer or more remote with a view to achieving this goal.

[26] Thus, among possibilities, there will be moral actions corresponding to the different moral duties, and different forms of implementation of these duties, and the morality of the *Gesinnung* will depend upon how close these moral possibilities are. The relevant maxims in a given situation of moral deliberation will therefore be those expressing the closest possibilities, while an act of incorporation will be required for agents to take some incentive as sufficient to determine her action. A very virtuous agent would only have moral possibilities in his vicinity. Here it is useful to understand the *Gesinnung* as a fundamental maxim of one's action (Allison 1990).

More specifically, to take an incentive as sufficient ground for action by incorporating it into one's maxim of action (Incorporation Thesis) can be re-phrased as taking this incentive's incorporation into the maxim to define that possibility which is the 'closest to me' as I understand myself in terms of my *Gesinnung*. In this way, the norms governing the actions of a SC are defined by a freely chosen ordering of possibilities involving her agency.

But one can also find grounds for interpreting Kantian spontaneity in terms of possibility by recalling that Kant's theory of objectivity rests upon the Transcendental Unity of Apperception as its highest principle. As Kant famously states, "The I think must be able to accompany all my representations" (B131). What is necessary is therefore the possibility of some act of spontaneity. So for Kant, objectivity is constituted by the possibility of a spontaneous 'I think'.[27] And this takes on an inter-subjective dimension because the 'I' could belong to any individual subject: what is actual is defined in terms of any possible spontaneity. Kant's Transcendental Idealism therefore reverses the priority of modalities characteristic of realism. For the realist, the actual is given, and the question of how to account for the possible then arises. For Kant, the possible defines the objective constraints on what is actual: this is actuality for any SC, that is, for any possible human perspective upon being, which perspective is characterized by being spatio-temporal and determined under the categories of the understanding. Here, the normative constraints upon the determination of the actual are therefore those defined by human transcendental subjectivity, in particular, featuring the principles of pure understanding.[28] This defines a first way in which the actuality of the world of appearances is defined in terms of the SC.

[27] This formulation hides some complexities. What Kant claims is that cognition of an object depends upon the possibility of accompanying one's representations with an 'I think' (i.e. the possibility of an act of spontaneity). So, representations contributing to my cognition are defined in terms of the necessity of this possibility. Insofar as an object is nothing but representations (whereby this is not to be understood as an identification of object and representation but as spelling out that objective determinations are nothing beyond the content of representations), this means that objectivity itself is defined in terms of the necessity of the possibility of an 'I think' accompanying these representations.

[28] In this determination of what is objective, the structuring of the possible is, of course, not up to the individual SC's freedom. Rather, it is a feature of any SC in virtue of its nature as transcendental subjectivity, i.e. of its having spatio-temporal intuition and apperception from which the categories can arguably be derived (Schulting 2012).

With this notion of objectivity in place, SC in its individuality (rather than in its representing one of the species 'human spontaneity') also determines actuality in a second sense: this is practical freedom. As we saw earlier, spontaneity can enable the causality of an intelligible character to be effective in the world of appearances. And, if acting on the strongest inclination, SC lets a certain natural causality take its course.[29] In the first case, the normativity of the moral law can be expressed in the world of experience, in accordance with the normativity of natural laws, while in the second case, it is only the latter normativity which is at stake. The choice of moral or non-moral action, and the further specifications of either reflect the agent's *Gesinnung*, and therefore the structuring of the domain of possibility that characterizes this particular spontaneity.

This structuring of the domain of possibility which defines the agent's *Gesinnung* can be viewed as part of a more general structuring of the possible by this particular SC which defines all the cognitive 'oughts' that Sartre refers to in his discussion of negation, as we saw above.[30] In a Kantian framework, these also have to cohere with those defined by theoretical spontaneity, while Sartre largely ignored the epistemological problem of accounting for the objectivity of our experience.[31]

The structuring of the possible by SC can thus include the constraints of Kant's practical and theoretical normativities. The closest possibility defines an 'ought' which reflects:

- the type of normativity defined by Sartre's examples of what I expect, for which, in a Kantian framework, the theoretical normativity of human spontaneity (ultimately, the Transcendental Unity of Apperception) provides a constraint defining what is objective; and
- when the possibility in question involves my agency, the practical norms defined by the fundamental project for Sartre, and by the

[29] As mentioned earlier, there is also action on other inclinations which involves a role for spontaneity, but I am leaving this out of the present account for the sake of simplification.

[30] This defines the subjective perspective on the actual which characterises each particular SC.

[31] Sartre's analysis of knowledge is of an ontological relation, and his discussion of empirical reality (BN: 180–216) does not seek to establish the necessity of any particular conditions for objective knowledge (Gardner 2008: 108).

> *Gesinnung* for Kant; in both cases, the agent's freedom is the ground of this normativity, but in Kant's case, this is defined in terms of the individual's degree of endorsement of the moral law as binding for all human agents.

This brief outline of a possible alternative account of the negation involved in the spontaneity of consciousness leads us to redefining SC in terms of modality. A SC is a structure in the domain of possibility. For all possibilities, this structure is imposed by our transcendental subjectivity (theoretical normativity). For those possibilities involving my agency, a free choice of my *Gesinnung* grounds my structuring of the possible (practical normativity involving some degree of endorsement of the normativity of the moral law). The form of non-being of SC is that of possibility.

7 Conclusion

In this chapter, I have sought to show that Kant's account of freedom raises ontological questions which, even if no knowledge of their answers is possible, require that there be a way of making sense of them which avoids contradiction, if the cogency of the notion of freedom is to be preserved. These ontological questions are also closely related to fundamental interpretative questions in Kant's ethics about the possibility of free immoral action. By drawing upon Sartre's ontology, his understanding of consciousness as lying beyond being In-Itself and of its relation to being In-Itself, and by showing the compatibility of aspects of Sartre and Kant's philosophical outlooks, a proposal has been formulated.

Our practical spontaneity can thus be understood as lying outside being (in-itself or appearances) while being related to it according to the agent's disposition (*Gesinnung*). This relation defines how practical spontaneity as *Willkür* enables the causality of the intelligible character to appear as further determinations of the agent's empirical character. Further, the chapter has outlined a proposal to define our spontaneous consciousness in terms of possibility, as a structure in the domain of the possible.

Although, in its final form, it goes beyond Sartre's ontology to give pride of place to modal distinctions, at its core, it relies upon an important linkage between Kant's transcendental conditions and Sartre's ontology of the For-Itself as that which happens to the In-Itself.

References

Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
—— (2004) *Kant's Transcendental Idealism. An Interpretation and Defense*. 2nd edition, New Haven: Yale University Press.
Baiasu, S. (2011) *Kant and Sartre. Re-discovering Critical Ethics*. Basingstoke: Palgrave Macmillan.
Baldwin, T. (1980) "The Original Choice in Kant and Sartre", *Proceedings of the Aristotelian Society*. 80(1): 31–44.
Davidson, D. (1963) "Actions, Reasons, and Causes", *Journal of Philosophy,* 60(23): 685–700.
Gardner, S. (2009) *Sartre's Being and Nothingness*. London: Routledge.
Heidegger, M. (1962) *Being and Time*. Tr. J. Macquarrie and E. Robinson. Oxford: Blackwell.
Kant, I. (1959) *Foundations of the Metaphysics of Morals*. Tr. L.W. Beck. Indianapolis, IN: Bobbs-Merrill.
—— (1991) *The Metaphysics of Morals*. Tr. M. Gregor. Cambridge: Cambridge University Press.
—— (1995) *Critique of Practical Reason*. Tr. L.W. Beck. New York: Macmillan.
—— (1997) *Critique of Pure Reason*. Tr. P. Guyer and A.W. Wood. Cambridge: Cambridge University Press.
—— (2009) *Religion within the Bounds of Bare Reason*. Tr. W.E. Pluhar. Indianapolis, IN; Cambridge, MS: Hackett.
Korsgaard, C. with Cohen, G.A., Geuss, R., Nagel, T. and Williams, B. (1996) *Sources of Normativity*. O. O'Neill – ed. Cambridge: Cambridge University Press.
Kosch, M. (2000) "Freedom and Immanence", J. Giles – ed. *Kierkegaard and Freedom*. London: Macmillan/St Martin's Press.
—— (2006) *Freedom and Reason in Kant, Schelling and Kirkegaard*. Oxford: Oxford University Press.
Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
O'Connor, D. (1985) "Good and Evil Dispositions", *Kant Studien,* 76: 288–302.
Onof, C. (1998) "A Framework for the Derivation and Reconstruction of the Categorical Imperative", *Kant Studien,* 89(3): 410–27.
—— (2010) "Kant's Conception of the Self as Subject and Its Embodiment", *Kant Yearbook,* 2: 147–74.
—— (2011) "Thinking the In-Itself and Its Relation to Appearances", in *Kant's Idealism. New Interpretations of a Controversial Doctrine*. Ed. D. Schulting and J. Verburgt. Dordrecht: Springer, 211–36.
—— (2011a) "Moral Worth and Inclinations in Kantian Ethics", *Kant Studies Online*, http://www.kantstudiesonline.net/KSO_Date_files/OnofChristian00711.pdf.
—— (2013) "Sartre's Understanding of the Self", in *Jean-Paul Sartre. Key Concepts*. Ed. S. Churchill and J. Reynolds. Durham: Acumen.
—— (2013a) "Review of Sorin Baiasu 'Kant and Sartre: Rediscovering Critical Ethics' ", *Kantian Review,* 18(2): 323–8.
Prauss, G. (1983) *Kant über Freiheit und Autonomie*. Frankfurt am Main: Klostermann.
Reinhold, K. L. (1790–2) *Briefe über die Kantische Philosophie*. Vol. II. Leipzig: Göschen.

Sartre, J.-P. (1958) *Being and Nothingness*. Tr. H.E. Barnes. London: Methuen.
—— (1970) *L'Existentialisme est un Humanisme*. Paris: Nagel.
Schelling, F. W. J. (1809) "Philosophische Untersuchungen über das Wesen der menschlichen Freiheitund die damit zusammenhängenden Gegenstände", *F. W. J. Schellings philosophische Schriften: Erster Band*. Landshut: Philipp Krüll.
Schulting, D. (2012) *Kant's Transcendental Deduction and Apperception. Explaining the Categories*. London: Palgrave Macmillan.
Wood, A. (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

Part III
Metaethics

6
Self-Knowledge in Kant and Sartre

Leslie Stevenson

I start with Sartre and make connections with Kant as we go along. In Section 1, I connect Sartre's "pre-reflective cogito" with Kant's "transcendental unity of apperception". In Section 2, I try to clarify Sartre's distinction between pure and impure reflection. In Section 3, I compare Kant on inner sense and apperception with Sartre on purifying reflection.

1 The pre-reflective *cogito*

To understand Sartre on self-knowledge we need to grapple with his challenging theory of human consciousness and the radical freedom that it involves, which is the unifying theme running through his early work. The first thing to notice is what he calls "the pre-reflective cogito", which he says is the necessary condition of the Cartesian cogito and any other sort of self-conscious reflection, for it is "the non-reflective consciousness which renders the reflection possible" (Sartre 1958[1943]: xxix). A mundane example he offers is that if one is counting the cigarettes in one's case, one's attention is on the cigarettes and the number of them, not on oneself; but if someone asks "What are you doing?" one can immediately reply "I'm counting my cigarettes". As he puts it, "The consciousness of man *in action* is non-reflective consciousness" (Sartre 1958[1943]: 36). In his short early study *The Transcendence of the Ego* (Sartre 1957[1937]), which makes a very helpful introduction to the dark depths of *Being and Nothingness*, Sartre offered similar examples of hanging a picture or repairing a tyre, where one can readily say what one is doing (or trying to do) if asked. Perhaps one does not need the stimulus of another's question in order to *think* "I am doing X", for if one gets distracted about what one is engaged in, one can ask oneself

"What am I trying to do here?" (though in short-term memory loss one may not be able to recall why one came into a room).

In his *Sketch for a Theory of the Emotions* (Sartre 1962[1939]), Sartre says that *emotional* consciousness is at first non-reflective, for it is primarily consciousness of the world, emotion being a specific manner of apprehending the world. For example, when one is angry with someone or at some situation, one is primarily aware of the qualities of the person or state of affairs that one sees as annoying or frustrating. The word 'angry' may not pass one's lips, and the concept of anger may not be in one's consciousness; a person's facial expressions, tone of voice and style of behaviour often reveal anger more clearly than what they say, and someone may even shout 'I'm *not angry*' in a way that unintentionally shows their anger. Yet some conscious awareness of our emotional states is possible: people sometimes say 'I'm angry with you', or 'I am angry with so-and-so about such-and-such', thereby displaying a certain level of reflection. However, Sartre makes some interesting further distinctions, as we will see in the next section.

The pre-reflective cogito also applies to sensations and thoughts. One can be asked what bodily sensations one presently feels, and one can usually give an immediate report of heat or cold, tiredness or pain, pressure or strain. One can also be asked 'What are you thinking about?' (or in the traditional intrusive phrase: 'A penny for your thoughts?'), and though there is usually no particular obligation to disclose one's mind, one can do so if one wishes, with a statement such as 'I was just thinking that p', 'I was remembering q', or 'I'm wondering whether r'. Some thoughts involve emotive attitudes ('I'm worrying about A', 'I feel annoyed with B about s', 'I'm delighted that t'), but Sartre has some challenging claims about emotion that we will soon get to.

Anyone who knows their Kant will point out that he seems to have anticipated Sartre's doctrine of the pre-reflective cogito when he wrote in his famous Transcendental Deduction:

> The *I think* must be *able* to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me. (Kant, 1998[1781/1787], B131–2)

In fact, Sartre started out from that famous saying of Kant's in *The Transcendence of the Ego* and was concerned to elucidate what it does and does not imply.

Neither philosopher makes an explicit contrast with animal mentality, but I suggest this would be illuminating. Animals perceive their world and react accordingly for survival and reproduction. The cat hears, sees and smells a mouse; she stalks it, catches and eats it; she sees another cat entering her territory and hisses and yowls, though if it is a tom and the time is right, she may invite coition. In one sense of the very ambiguous notion of consciousness, our moggy may be said to be *conscious* of her mouse, her rival, and her mate. But of course she cannot *say* what she is perceiving or what she is doing, and presumably she cannot *think* it either, for she cannot be credited with the relevant concepts; the 'I think' cannot accompany her representations. So if we are to speak of animal consciousness at all, it has to be understood in a different sense from the human pre-reflective cogito, let alone the reflective level. The apparent ability of chimpanzees or dolphins to recognize themselves in mirrors raises interesting questions about what sense of self we can credit them with, and there are related questions about human infants before they learn to use the first-person pronoun; but the difference from normal human adults remains clear, because of the use of language.

In his short paper "Intentionality" (Sartre 1970[1939]), Sartre asserts the fundamental intentionality of human consciousness and rejects what he memorably calls "the digestive philosophy" according to which the subject reduces objects to perceptual appearances, like a spider wrapping its prey with its own silk before eating it. Sartre follows the lead of Husserl in maintaining that consciousness is always consciousness *of* something. He talks of "a transcendent object", but he does not mean a denizen of a metaphysical realm transcending the senses, only anything thought of as existing independently of the subject's consciousness. In the case of mirages or hallucinations, there may be no actual object answering to the content of the subject's mental state, but her consciousness still has an "intentional object", i.e. she *thinks* there is an object of the relevant sort out there in the world. Sartre must have been thinking of *positional* consciousness, as opposed to the non-conceptual awareness characteristic of animals and infants. But, like Kant, he was not very careful about his terminology: though Sartre occasionally talks of "non-positional consciousness", he would need to rephrase that as non-positional *awareness* or *perception* if he is to maintain his thesis that all consciousness involves the positing of an object.

It would be a misleading interpretation of Sartre's pre-reflective cogito to say that all consciousness (in normal human adults) involves a second-order consciousness of consciousness. Although he says at one point "the necessary and sufficient condition for a knowing consciousness to

be knowledge of its object is that it be consciousness of itself as being that knowledge" (Sartre 1958[1943]: xxviii), he insists that this does *not* mean that to know is to know that one knows. Such positional awareness of one's own mental states is involved in reflection, but not in our usual states of pre-reflective consciousness. One of Sartre's most precise statements of the pre-reflective cogito is that "every positional consciousness of an object is at the same time a non-positional consciousness of itself" (Sartre 1958[1943]: xxix). He also remarks that in ordinary actions like counting one's cigarettes, one has a positional consciousness of the objects of one's attention but only a "non-thetic" consciousness of one's activity. Frustratingly, Sartre does not define these technical terms, but Sebastian Gardner in his *Reader's Guide* to *Being and Nothingness* helpfully explains that *positional* consciousness is consciousness of an intentional object O and *thetic* consciousness is consciousness involving a proposition that p (Gardner 2009: 45).

Should we say then that the cat is positionally conscious of the mouse, but not thetically conscious of it, since the cat cannot entertain any propositions about it? That will depend on whether positional consciousness of an object is defined as involving a *concept* of it. The issue also arises in the interpretation of Kant in the question whether Kantian "intuitions" (*Anschauungen*, perceptual or quasi-perceptual states) are conceptualized or non-conceptual. His fundamental distinction between sensibility and understanding and thereby between intuitions and concepts (Kant 1998[1781/1787], A19/B33 and A51/B75) strongly suggests that intuitions must be *non*-conceptual. But when he treats space and time as "intuitions", he gets into notorious obscurities when he fails systematically to distinguish space and time, *perceptions* of space and time (or spatial and temporal relationships), and *concepts* of space and time. In the *Jäsche Logic*, Kant did distinguish unconceptualized and conceptualized perceptions when he mentioned the possibility of a "savage" perceiving a house without knowing that it is a house. He left the topic in a bit of a mess, but in his erratic genius he left us clues as to how to clear up the mess.

I suggest that whereas animals have only unconceptualized perceptions, humans can have *both* kinds, for as well as lots of conceptualized perception one can be non-conceptually aware of the movement of some unidentified object zooming towards one's head, or of a nameless noise or smell or pressure on the skin, and perhaps of the mood expressed on someone's face. In unreflective states of consciousness we are aware (conceptually or non-conceptually, positionally or non-positionally) of objects, and we are not thinking about ourselves. But

as soon as a question is asked of us as to what we are perceiving or doing (or *think* we are perceiving or doing), we are transported or transposed to the reflective level, and we become explicitly "positionally" aware of ourselves and explicitly "thetically" aware of facts about ourselves.

There is another danger to be steered around here. Saying that we can be positionally aware of ourselves may suggest that we are thus conscious of ourselves as a particular item within our experience, in the way that we can be conscious of a fly in our ointment or an intruder in our house. It is true that we can perceive parts of our own *bodies* as physical objects – I can see a smudge of dirt on my hand, I can feel a scar on my back, and I can realize that the weight of my whole body is bending the branch on which I am sitting. But when one says or thinks 'I'm trying to do X', 'I'm wondering whether p', 'I'm annoyed about q' that is not on the basis of a perceptual identification of one's own body, though in the case of *other* people's mental states one does of course have to rely on what they perceptibly do and say. In that sense, the self or 'ego' or 'I' is not one item among others in the contents of introspection. The point is a familiar one in philosophy. Hume made it with famous rhetorical force when he wrote:

> For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I can never catch *myself* at any time without a perception, and never can observe anything but the perception. (Hume 1978[1739]: I.IV.vi)

Kant made the point with more words but less memorability in his Transcendental Deduction and in the Paralogisms, where he deconstructs the argument that the self or subject of experience must be a substance:

> The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called inner sense or empirical apperception. (Kant 1998[1781/1787]: A107)

> Thus if that concept, by means of the term 'substance', is to indicate an object that can be given, and if it is to become a cognition, then it must be grounded on a persisting intuition … But now we have in inner intuition nothing at all that persists, for the I is only the

> consciousness of my thinking ... and the simplicity of substance that is bound up with the objective reality of this concept completely falls away and is transformed into a merely logically qualitative unity of self-consciousness in thinking in general. (Kant 1998[1781/1787]: B412–13; cf. B407, B421–2)

Sartre gave his own version of this negative thesis:

> When I run after a streetcar, when I look at the time, when I am absorbed in contemplating a portrait, there is no *I*. There is consciousness of *the-streetcar-having-to-be-overtaken*, etc., and non-positional consciousness of consciousness. ... There is no place for *me* on this level. And this is not a matter of chance, due to a momentary lapse of attention, but happens because of the very structure of consciousness.
>
> The *I* is not given as a concrete moment, a perishable structure of my actual consciousness. On the contrary, it affirms its permanence beyond this consciousness and all consciousnesses, and – although it scarcely resembles a mathematical truth – its type of existence comes much nearer to that of eternal truths than to that of consciousness. (Sartre 1957[1937]: 49–50)

Sartre's main business there was to argue against Husserl's positing of a "transcendent ego" as the subject and organizing principle of states of consciousness. Sartre wrote "the flux of consciousness constitutes itself as the unity of itself" (Sartre 1957[1937]: 60), which sounds like his version of Hume's "bundle" theory of personal identity. (I do wonder whether if Husserl had digested the lessons of Hume and Kant, a lot of phenomenological head-scratching and pen-scratching might have been saved!)

There is some unclarity, however, about exactly where the distinction between unreflective and reflective consciousness is supposed to fall. Does *any* statement or thought involving the first-person pronoun or concept automatically put one onto the reflective level? That is the most straightforward thing to say. Sartre talks of "any reflective statement: I read, I dream, I perceive, I act" (Sartre 1958[1943]: 156), and thus seems to follow this line. But earlier he wrote that the "I" can appear on the *un*reflected level, since one can say things like "I am trying to hang this picture" or "I am repairing the rear tyre" while remaining totally preoccupied with one's activity, thinking about what needs to be done

without thinking of oneself as doing it (Sartre 1957[1937]: 89). Other more emotive examples suggest the opposite, however. If one is absent-mindedly picking one's nose and is made aware of it by a comment or a nudge or a knowing look, it seems hard to deny that one is thereby transformed into a reflective, self-conscious mental state. Sartre's famous example of the person caught spying through a keyhole (Sartre 1958[1943]: 259–60) points in the same direction. Even in the mundane picture-hanging case, if one asks for help, saying "You hold the picture while I get the wire over the hook", that involves attention to what one is doing oneself as well as what the other is doing. It seems simplest to rule that as soon as the "I think" or "I am doing" accompanies a mental state, that can count as reflection. I suggest this as conceptual legislation rather than statement of empirical fact, for I doubt if the ordinary usage of the term 'reflection' will settle the matter. However, Sartre distinguishes different kinds of reflection, as we will now see.

2 Pure and impure reflection

Sartre's distinction between pure and impure reflection plays an important role in some crucial places in *Being and Nothingness*, but readers may well wish he had given us a more systematic explanation of it. Francis Jeanson put considerable emphasis on this distinction in his early exposition *Sartre and the Problem of Morality* (Jeanson 1980[1947]), which received the master's *imprimatur*, and Sebastian Gardner elaborates on it at several places in his recent *Reader's Guide to Being and Nothingness* (Gardner 2009). Sartre first introduced his pure/impure distinction in *The Transcendence of the Ego*:

> Pure reflection ... keeps to the given without setting up claims for the future. This can be seen when someone, after having said in anger, "I detest you", catches himself and says, "It is not true, I do not detest you, I said that in anger." We see here two reflections: the one, impure and conniving, which effects there and then a passage to the infinite, and which through the *Erlebnis* [i.e. the momentary experience] abruptly constitutes hatred as its transcendent object; the other, pure, merely descriptive, which disarms the unreflected consciousness by granting its instantaneousness. (Sartre 1957[1937]: 64–5)

So in impure reflection one attributes to oneself what Sartre calls "a state", a relatively long-lasting attitude or disposition (e.g. hating or

detesting someone) which in that sense "transcends" what is presently "given" to one's consciousness (though Sartre's talk of "a passage to the infinite" here seems to be inflated rhetoric). By contrast, pure reflection affirms only what the subject is conscious of in herself at the time. There is an obvious valuational connotation in the labels 'impure' or 'conniving', and there are stronger indications of this in *Being and Nothingness*. Sartre argues (Sartre 1957[1937]: 60–71, and in greater depth in Sartre 1958[1943]: IV.1.i) that conventional psychological explanations of actions in terms of beliefs, desires and other "psychic" states or "qualities" fail properly to explain, and that impure reflection involves a kind of self-deception or "bad faith".

In *Sketch for a Theory of the Emotions*, Sartre said that liberation from the illusion or "magic" of emotion in which one irrationally projects qualities onto the world "can come only from a purifying reflection or from the total disappearance of the emotional situation" (Sartre 1962[1939]: 81; see also Sartre 1958[1943]: 445). In impure or "accessory" reflection one says "I am angry because he is hateful", but pure reflection may enable one to say, with more insight, "I find him hateful because I am angry" (Sartre 1962[1939]: 91). The same would apply to more positive emotions: the lover may say 'I love you because you are so beautiful', when purifying reflection might suggest 'I find you beautiful because I'm in love with you' – which, however, might be in danger of destroying the magic! But that is Sartre's point: that emotion involves a sort of magical thinking, in which one "invests" the world of other people and objects with qualities that they may not really possess. Even the conventional words 'I love you' are dangerous, not just because the declaration may not be reciprocated, but because those words are ambiguous between what Sartre would call a pure reflective statement of the lover's present state of consciousness and an impure assertion that he will go on feeling the same for unspecified tracts of the future. The traditional wedding vow "to love you until death do us part" is a different matter, for that is not a promise of future *feelings* (though of course everyone hopes they will mostly fall into line) but rather a commitment to *act* in the appropriate ways. Such a promise is not a prediction but an explicit commitment to try to fulfil the condition, and it is possible to make such commitments without any "impure" projection of one's present feelings into the future. The addition of the pious phrase 'so help me God' implies recognition that one's own feelings cannot always be relied upon.

In *Being and Nothingness* the distinction between pure and impure reflection surfaces in the middle of the very obscure chapter on temporality in Part 2, where Sartre writes:

> Pure reflection, the simple presence of the reflective for-itself to the for-itself reflected-on, is at once the original form of reflection and its ideal form; it is that on whose foundation impure reflection appears, it is also that which is never first *given*; and it is that which must be won by a sort of catharsis. (Sartre 1958[1943]: 155)
>
> ...it is impure reflection which constitutes the succession of psychic facts or psyche. What is given first in daily life is impure or constituent reflection although this includes pure reflection as part of its structure. (Sartre 1958[1943]: 159–60)

There may seem to be a tension between saying on the one hand that pure reflection is "the original form" but on the other that it is "not given first", but I take Sartre's point to be that in the usual course of our mental lives we attribute persisting mental states and dispositions to ourselves ("psychic facts"), and that it takes a special kind of conceptual and mental hygiene ("catharsis") to make us realize that such attributions not only go beyond the evidence of our instantaneous self-consciousness but serve to hide from ourselves our fundamental freedom to change our mental attitudes, and hence our responsibility for them. Pure reflection is mentioned again in connection with catharsis in the final sentences, where Sartre promised another work "on the ethical plane", which never actually appeared (Sartre 1958[1943]: 628).

But what are we to say of that simplest kind of reflection that arises immediately out of the pre-reflective cogito, when one is asked what one is doing (or trying to do) and one replies e.g. 'I'm repairing the rear tyre'? Is that pure or impure? It does not seem to involve any sort of strenuous "catharsis", yet it is "pure" in the sense that it does not go beyond the contents of the agent's consciousness, knowledge without observation of his own present intentions. Perhaps Sartre needed to make a distinction *within* pure reflection between the simple immediate kind just mentioned and the "*purifying*" kind that involves reflection on one's deeper purposes and either self-conscious endorsement of them or a radical conversion to different ones. But so far as I know he does not make any such terminological distinction.

Sartre says that "it is in anguish that man gets the consciousness of his freedom" (Sartre 1958[1943]: 29), and he defines anguish as "the reflective apprehension of freedom by itself" (Sartre 1958[1943]: 39). That must surely be *pure* and indeed *purifying* reflection (though he does not say so), for he holds that though freedom is "a permanent structure of the human being", anguish is "completely exceptional". Most of the time we *flee* anguish, we avoid explicit "pure" consciousness of our inescapable freedom (Sartre 1958[1943]: 440) and conduct our lives in "bad faith". Early on Sartre makes a connection with values:

> It follows that my freedom is the unique foundation of values and that *nothing*, absolutely nothing, justifies me in adopting this or that particular value, this or that particular scale of values. As a being by whom values exist, I am unjustifiable. My freedom is anguished at being the foundation of values while itself without foundation. (Sartre 1958[1943]: 38)

Analytical philosophers can tidy up the rhetoric here, and say that it is individual human beings, not their "freedom" in the abstract, who are capable of reflective apprehension of their own free choices, and thereby of "anguish" in this existentialist sense of the word, which has roots in Kierkegaard and Heidegger. But whether Sartre intends or needs to be committed to the extreme subjectivism about values that is so strongly suggested in that passage is an important issue that lies outside the scope of this chapter.

3 Self-knowledge and freedom

Let me now try to clarify matters by raising the general question of whether, and under what conditions, there is such a thing as *self*-knowledge. It is uncontroversial that one can have knowledge of one's own bodily states and one's public history, including many of one's actions. I know that I was born on such-and-such a date, that my body bears the scars of an operation, and that I have written and rewritten this chapter: all that is what Sartre refers to as "my facticity". Other people can know the same facts about me, though there may be some things that I remember doing that nobody else has any record of. Memory is a special kind of self-knowledge, but in the cases we have just been considering the facts remembered are publicly available facts. In proprioceptive perception, I know, without depending on sight or touch,

the positions and flexings of my own limbs – which are perfectly objective physical facts, observable and measurable by others – but I know them in a *way* that is not available to others. Proprioception is a form of perception (with a very limited range), and it is made possible by internal feedback mechanisms in the body. It can thus be called a "sixth sense" or a literally "inner" or "internal" sense.

But all that is by way of contrast to what we are primarily interested in here, namely knowledge of one's own *mental* states. In the history of philosophy there has been a tendency to assume that mental states are in their own peculiar way just as objective as physical states, being part of the total reality of the universe, though there is a uniquely first-person way of knowing them. Descartes argued that our knowledge of our own conscious states is infallible, in contrast to the fallibility of all outer perception. Locke presented "reflection", the perception of the operations of our minds within us, as the source of all our ideas of mental states, alongside "sensation" for our ideas of outer objects, and he said "though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense" (Locke 1975[1690]: II.i.4). Hume followed Locke when he talked of impressions of reflection as well as sensation (Hume 1978[1739]: I.i.2). As we have seen above, Hume denied that there is a perception of one*self*, but he affirmed that one is always aware of some *particular perception* or other (where for Hume the word 'perception' includes any sort of mental state or event): he gave a varied set of examples including sensations of heat or cold, perceptions (in the modern sense) of light or shade, emotional attitudes of love or hatred, the sensation of pain, and the feeling of pleasure.

Kant at first seems to follow the lead of Locke and Hume when he distinguishes outer sense – by which he means perception of physical objects in space – from inner sense "by means of which the mind intuits itself, or its inner state" (Kant 1998[1781/1787]: A22–3/B37). But as we have already noted, he later argues very forcefully that there is no intuition of the self, and indeed he immediately qualifies the line just quoted by adding that inner sense gives no intuition of the soul itself, which leaves just the claim that by inner sense the mind intuits its inner states. To translate out of Kant-speak: we have a quasi-perceptual awareness of our own mental states (or at least of *some* of them, because Kant anticipates Freud and cognitive science in allowing the possibility that some of our mental states may remain unconscious). Later in the first *Critique* he emphatically distinguishes this inner sense or "empirical

apperception" from "transcendental apperception", namely the principle (noted above) that the "I think" must be able to accompany any of my representations (see Kant 1998[1781/1787]: A107 and B153).

But even the modest claim that we have a quasi-perceptual awareness of some of our own mental states comes under some pressure in the development of Kant's thought (Guyer and Wood note that Kant continued to worry about the problem of inner sense until the end of his career – see Kant 1998[1781/1787]: 727 n 43). In sections 24 and 25 of the Transcendental Deduction (one of the most difficult passages in one of the most difficult chapters in all philosophy!), Kant wrestles with the notion of self-knowledge through inner sense. I do not have the time and space to expound systematically what he argues there, but I will try to grasp the main thread. He finds paradoxical the implication of his conception of inner sense that we can know ourselves only as we appear to ourselves, not as we are in ourselves. For if inner sense is really a kind of *sensibility*, its manifold of data (like those of outer sense) have to be combined or synthesized by our faculty of understanding if we are to bring the raw unconceptualized data under concepts and make judgements about ourselves. So Kant reaffirms that in inner intuition "I have no cognition of myself as I am, but only as I appear to myself" (Kant 1998[1781/1787]: B158), but he goes on to distinguish between inner sense and the original unity of apperception (i.e. transcendental apperception), in which, he says:

> I am conscious of myself not as I appear to myself, nor as I am in myself, but only *that I am*. This representation is a thinking, not an intuiting. (Kant 1998[1781/1787]: B157, with my emphasis)

At this point Kant brings in his belief in the "spontaneity" or "self-activity" of conceptual thought, as opposed to the passive "receptivity" of sensibility:

> The combination of a manifold in general can never come to us through the senses ... for it is an act of the spontaneity of the power of representation, and, since one must call the latter understanding ... all combination ... is an action of the understanding which we would designate with the general title synthesis. (Kant, 1998[1781/1787]: B129–30; cf. A97, A126–7)

And in the footnote at B157–8 he says that "the I think expresses the act of determining my existence", which presumably follows from his

claim that in the original unity of apperception I am conscious "that I am". So transcendental apperception yields knowledge only of my bare existence, not of any of my mental states; and yet Kant says it does involve knowledge of my own spontaneity, for in it I "merely represent the spontaneity of my thought", and "this spontaneity is the reason I call myself an intelligence".

There is an obvious cue here for comparison with Sartre's conception of our fundamental spontaneity or freedom – the unstable, non-substantial, forever-to-be decided nature of human consciousness. But let us stay with Kant just a little longer, for he had more than one conception of spontaneity. In the first *Critique*, he argues that our judgements involve concepts as well as intuitions, so our judgements and beliefs are not determined by our sensory input alone. But in his moral philosophy Kant lays enormous stress on the difference between two kinds of motives for action, namely our natural inclinations and our recognition of moral duties. "Sensibility" in the epistemological context refers to the stimulations of our sense-organs, whereas "sensibility" in the moral context refers to self-interested desires. Correspondingly there are two kinds of *reasons* – for beliefs and judgements, and for desires, intentions and actions.

In his practical philosophy, Kant acknowledges deep difficulties in knowing our own reasons for action, because our motives are typically mixed, so it can be hard to be confident about our real motivation. In the first *Critique* he declares that "the real morality of our actions (their merit and guilt) even that of our own conduct, therefore remains entirely hidden from us" (Kant 1998[1781/1787]: A552/B580 footnote). In the *Groundwork* we find the following eloquent statement (well, quite eloquent for Kant!):

> In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty. It is indeed sometimes the case that with the keenest self-examination we find nothing besides the moral ground of duty that could have been powerful enough to move us to this or that good action and to so great a sacrifice; but from this it cannot be inferred with certainty that no covert impulse of self-love, under the mere pretence of that idea, was not actually the determining cause of the will; for we like to flatter ourselves by falsely attributing to ourselves a nobler motive. (Kant 1996[1785]: 4:407)

In later works, Kant develops this theme of the opacity of our motives to ourselves, notably in the Method section of the *Critique of Practical Reason* (Kant 1996[1788]5: 151–63), and in *The Metaphysics of Morals* (Kant 1996[1797]6: 446–7) where he says quite definitely that "the depths of the human heart are unfathomable". Yet despite this difficulty, Kant presents self-knowledge as "the First Command of All Duties to Oneself":

> This command is "know (scrutinize, fathom) yourself", not in terms of your natural perfection ... but rather in terms of your moral perfection in relation to your duty. That is, know your heart – whether it is good or evil, whether the source of your actions is pure or impure.
>
> Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one's heart which are quite difficult to fathom, is the beginning of all human wisdom. (Kant 1996[1797]: 6:441)

It seems that this Socratic and Kantian injunction to know oneself can only be presented as a regulative ideal, to know oneself as far as possible. Sartre claims that we live most of our lives in "bad faith", not clearly or reflectively aware of our own motives, and he says that "what we might call everyday morality is exclusive of ethical anguish". But he wants to insist that bad faith is not inevitable, that we can face up to ethical anguish, perhaps in response to Socratic questioning, and use our potential for purifying reflection.

There may seem to be a difference between Sartre and Kant here, in that where Kant talks of self-*knowledge* (if only as an ideal), Sartre's language suggests that purifying reflection is more a matter of *deciding* with full self-conscious clarity what values or ends one is going to pursue. However, the apparent gap may be closed when we remember that on the one hand Kant is not only concerned with theoretical self-knowledge but with living up to the demands of the moral law. He would obviously not be impressed by a clear-headed tyrant or gangster or terrorist or child abuser who knows his own motivations as clearly as anyone does, but would remind such a person that there are moral decisions to be made, and that he remains fundamentally free to give up the evil and choose the good. And on the other hand Sartre will want to say that in leading up to such moments of moral decision or "radical conversion" there will need to be a clear-headed assessment of one's actions and motivations up to that point (the "facticity of one's past"), as part of the process of purifying reflection.

Questions can also be raised on the theoretical side about judgement and belief. At the beginning of Part 2 of *Being and Nothingness* (another of the most difficult passages in all philosophy), Sartre says:

> The being of consciousness is a being such that in its being, its being is in question. This means that the being of consciousness does not coincide with itself in a full equivalence. (Sartre 1958[1943]: 74)

An even more pithy version of this is his apparently contradictory formula "consciousness is not what it is" that recurs throughout the book. To explore its full meaning for Sartre would need another paper, but perhaps we can make some sense of its application to belief, when he says "To believe is to know that one believes, and to know that one believes is no longer to believe" (Sartre 1958[1943]: 69). He admits that that formulation has "forced the description" by using the word 'know', so presumably his point is first, that the pre-reflective cogito applies to belief, so one can become self-consciously aware of one's beliefs; but second, that when one becomes reflectively conscious that one has a certain belief, that raises the question of what *reasons* one has for it, and hence of whether to *maintain* that belief:

> Thus by the sole fact that my belief is apprehended as belief, it is *no longer only belief*: that is, it is already no longer belief, it is troubled belief. (Sartre 1958[1943]: 74–5)

A similar point applies on the practical side:

> By the sole fact that I am conscious of the reasons (*motifs*) which inspire my action, these reasons are already transcendent objects of my consciousness; they are outside. In vain shall I try to catch them; I escape them by my very existence. I am condemned to exist forever beyond my essence, beyond the reasons and motives of my act. I am condemned to be free. (Sartre 1958[1943]: 439)

In terms of Sartre's example at 443, a soldier's desire to save his own life could be involved *both* in the instinctive reaction of running away from attack and in a more reflective decision to stay at his post because he reckoned that was the most likely way of surviving, whereas staying put might rather be motivated by defending his country even at risk to his life. For Sartre, most deliberation involves only impure reflection about

which *means* to adopt to the desires one already has; he says "When I deliberate, the chips are down" (*les jeux sont faits*) (Sartre 1958[1943]: 450–1). But he implies that there is a more fundamental kind of free choice that involves purifying reflection about which desires to endorse and act on, which *ends* to adopt.

By the pre-reflective cogito or empirical apperception one can be consciously aware of various beliefs, desires and emotions in oneself at a given time, and in pure reflection we can acknowledge them, perhaps publicly – but as we have seen in Section 1, only as temporary mental states. What we "really" believe or want or feel is a deeper question which requires not just pure but *purifying* reflection, and is not so much a matter of finding out some obscure fact (in the Freudian unconscious, perhaps) but of *deciding* in the light of whatever we think best. In the spirit of the existentialism that is explicit in Sartre and implicit in Kant, let me end with a real-life example: "Do you really *believe* that she will stay with you, after all that has happened?" and "Do you really *want* her to stay with you, all things considered?"

I think I have shown that there are deep connections between Kant's philosophy of mind and Sartre's, deeper indeed than Sartre himself seems to have realized, and that their shared insights are still largely valid. Of course, there is an obvious tension between Kant's fundamentally rational and objective approach to ethics and the radical subjectivity of all value that is suggested by the rhetoric of Sartre's early philosophy – which is all that I have considered here. In later work he tried to row in a different direction, but that is another story.

References

Gardner, S. (2009) *Sartre's Being and Nothingness: A Reader's Guide*. London: Continuum.

Hume, D. (1978[1739]) *A Treatise of Human Nature*. Ed. P. H. Nidditch. Oxford: Oxford University Press.

Jeanson, F. (1980[1947]) *Sartre and the Problem of Morality*. Tr. R. Stone. Bloomington: Indiana University Press.

Kant, I. (1996[1785]) *Groundwork of the Metaphysics of Morals*. Tr. M. Gregor. In *Practical Philosophy*. Cambridge: Cambridge University Press.

—— (1996[1788]) *Critique of Practical Reason*. Tr. M. Gregor. In *Practical Philosophy*. Cambridge. Cambridge University Press.

—— (1996[1797]) *The Metaphysics of Morals*. Tr. M. Gregor. In *Practical Philosophy*. Cambridge: Cambridge University Press.

—— (1998[1781/1787]). *Critique of Pure Reason*. Tr. P. Guyer and A. Wood. Cambridge: Cambridge University Press.

Locke, J. (1975[1690]) *Essay Concerning Human Understanding*. Ed. P. H. Nidditch. Oxford: Oxford University Press.
Sartre, J.-P. (1957[1937]) *The Transcendence of the Ego*. Tr. F. Williams and R. Kirkpatrick. New York: Farrar, Straus and Giroux.
—— (1958[1943]) *Being and Nothingness*. Tr. H. Barnes. London: Methuen.
—— (1962[1939]) *Sketch for a Theory of the Emotions*. Tr. P. Mairet. London: Methuen.
—— (1970[1939]) "Intentionality: A Fundamental Idea of Husserl's Phenomenology". Tr. J. Fell. *Journal of the British Society for Phenomenology,* 1(2): 4–5.

7

Action, Value, and Autonomy: A Quasi-Sartrean View

Peter Poellner

1 Outline of a quasi-Sartrean theory of action

In this chapter I shall examine one of the main contributions of the phenomenological tradition to practical philosophy and the theory of value: Jean-Paul Sartre's. Some of the issues it addresses also occupy centre stage in current debates on practical rationality. I hope to show that Sartrean phenomenology has something both distinctive and plausible to contribute to these discussions. The subtlety and complexity of the early Sartre account have often gone unrecognized or under-interpreted, partly as a result of misreadings, some of which, to be sure, have been motivated by ambiguities in his own utterances. As we shall see, his conclusions about practical rationality and value have affinities with ideas found in other approaches, particularly in those tracing their origins to Scheler and Kant. Since these are usually considered to be in competition with each other, Sartre could be seen as delineating a rather surprising synthesis. My reflections in this chapter belong to a genre in the history of philosophy perhaps best described as rational reconstruction. In this spirit, I shall look at a number of Sartrean theses, taken from the following list of claims which, I submit, jointly make up the core of his practical philosophy:

(1) Human action is in large measure intentional action, and (at least) most intentional action is responsive to apparent reasons that seem to speak in favour of the action (justifying reasons) (Sartre 2003: 455–8).[1]

[1] Sartre's writings are cited from the translations listed in the bibliography, which I have sometimes modified.

(2) Presumptively reasons-responsive action is essentially conscious. It is our phenomenal *awareness* of apparent reasons that enables them to guide some of our behaviour in the way that is necessary for that behaviour to be engaged in *for* those reasons (Sartre 2003: 475).
(3) Our consciousness of apparent justifying reasons is grounded in a consciousness of values (Sartre 2003: 62, 117–19).
(4) Some of the values guiding and potentially rationalizing intentional action are encountered in the empirically objective world (Sartre 1999: 50–1; 2002a: 383–4; 2002b: 57–8; 2003: 62; 2004b: 17–20).
(5) The fundamental, direct mode of access to values is through conscious affectivity, which is therefore, in successful cases, a mode of *acquaintance* with value (Sartre 2003: 57, 62, 465).
(6) Where these values qualify empirical objects, conscious affectivity often has the structure of *perception* (Sartre 2002b: 55–8).
(7) Not all value is empirically object-qualifying. The justifying force or authority of the values of objects is overridden by the value of reasons-responsive consciousness as such (Sartre 1980: 51–2; 1999: 96, 108).
(8) Insofar as consciousness is adequately reasons-sensitive, it is therefore fundamentally guided by normative demands deriving from the value of reasons-responsive consciousness itself (Sartre 1980: 51–2).
(9) If consciousness is motivated by such normative demands, it can be said to *determine itself* and to be, in this sense, autonomous (Sartre 1992: 478, 481–2).
(10) The positive value of self-transparent ('authentic') and adequately reasons-sensitive autonomous consciousness is 'absolute': it is in each instance greater than the aggregated value of anything else that is not itself such a consciousness or capable of becoming one (Sartre 1999: 96).

A further thought that I want to attribute to Sartre, although this would be disputed by some interpreters, is that:

(11) All the above claims are independent of the *metaphysical* status of the items referred to. They concern the practical reality of action and practical deliberation – i.e. they are about the, for us, inescapable phenomenological 'life-world' – and they are true, if they are true, irrespective of whether phenomenal consciousness or values belong among the ultimate furniture of the universe (Sartre 1980: 56; 2003: 57–8).

There is, finally, the following thesis that Sartre is usually also taken to hold:

(12) A conscious agent is fully autonomous even when she is not adequately sensitive to genuine reasons but is guided by merely apparent reasons grounded in whatever appear to the agent to be valuable ends at the time of action (Sartre 2003: 49, 464–7).

There are many statements by Sartre, especially in *Being and Nothingness*, that support attributing this thesis to him. He sometimes suggests, for example, that the being of a value depends on any individual agent's actual commitment (Sartre 2003: 62), which seems to make the grounds of reasons, and therefore the reasons themselves, agent-relative, indeed, relative to agents-at-a-time. He occasionally suggests, further, that an agent may act fully autonomously even when she systematically adopts means in the pursuit of her ends that are manifestly inappropriate to the attainment of those ends (2002b: 41–2; 2003: 467). And he asserts that the ultimate ends pursued at least by non-self-transparent, i.e. inauthentic, agents are incoherent (although the agents do not know this), being determinate versions of the determinable end to be an in-itself-for-itself, a fully self-sufficient consciousness that is also, qua lived consciousness, an object (Sartre 2003: 114–15).[2] Despite this, such agents are autonomous, according to Sartre. Given his further claim that the fundamental apparent reasons guiding free actions are based on values that the agent associates with the ends of those actions, this entails that the fundamental putative reasons of free action are incoherent and hence can never be genuine or good reasons (i.e. they cannot be reasons, period).

If this is Sartre's view, one might justly think that his conception of autonomy is so far removed from more standard conceptions of it – however fuzzy and contested these may be in various respects – as to amount to a changing of subject matter. He might then be thought, not implausibly, to have conflated *intelligible* action – action that makes sense from the agent's perspective at the time – with *autonomous* action. To have explicated the structure of intelligible action would be no small thing, but it is not quite what Sartre has aspired to doing.

However, his position on the issues just mentioned is not unambiguous and his various utterances on them are not always consistent.

[2] This is an incoherent end because lived consciousness in its fundamental mode of self-presence cannot be given to itself as an object. See footnote 4.

Only two years after the publication of *Being and Nothingness*, he writes, apparently in clarification and without taking himself to be retracting anything he has said there, that goods (values), while existentially dependent on consciousness, are 'universal', agent-neutral noematic contents[3] that are not dependent on any one individual consciousness (Sartre 1992: 555–7; 1980: 29–30), nor created or conferred by individual choice or commitment (Sartre 1992: 517–18). It is very plausible to extend this clarification also to instrumental reasons, that is, to considerations concerning appropriate means for achieving an action's ends, although Sartre himself does not explicitly do so. This leaves, from the obviously problematic claims about the conditions sufficient for autonomy mentioned in the preceding paragraph, only the idea that an agents can be fully autonomous even if she is oriented towards ultimate ends that are demonstrably irrational. This is an idea that is deeply entrenched in Sartre's overall position and cannot plausibly be interpreted away. But it is also an idea in support of which he offers neither good arguments nor convincing phenomenological descriptions. Why should we accept that the ultimate determinable end of all inauthentic human agency is *being an ontologically self-sufficient conscious object*? If we look at Sartre's explicit arguments, we find that they are mostly arguments for a different claim: human consciousness, insofar as it is projective (end-pursuing) is necessarily characterized by an experienced lack or deficiency, experiencing itself as unfulfilled or 'incomplete'; every action involves, usually implicitly, a conscious desire to overcome that lack or deficiency and to become a 'completed' consciousness, one that would not experience itself as lacking anything (Sartre 2003: 111–13). I shall call this Sartre's *completion thesis*. It is an idea worth debating and it has a long philosophical ancestry. The idea of such an ultimate end is not *manifestly* irrational. But Sartre, without much discernible argument, takes it to be equivalent to the ontological aspiration to become a

[3] This point of Sartre's needs to be qualified if it is to be consistent with his claims about the value of consciousness (see Section 3). The noematic content of an intentional experience is an intentional-object-involving content; it consists in the intentional object considered just as it is given in the experience. If *experiencing* itself, as Sartre argues, is phenomenologically distinct from *what* is experienced; and if experiencing is, while necessarily consciously present to itself, fundamentally not an intentional object for itself; and if it has intrinsic value; then that value cannot be fundamentally presented by way of a noematic content. What Sartre should more precisely have said is that every actual value, while not necessarily itself being an agent-neutral noematic content, can in principle and without evaluative distortion be indirectly presented via some such content. For more details, see Poellner (forthcoming).

God-like consciousness and maintains that such a consciousness would have to have the mode of being of an object, in the broad phenomenological sense of this expression, making the idea of it self-contradictory if one accepts his premise that consciousness as lived is necessarily aware of itself but cannot in principle be fundamentally aware of itself as an object.[4] I submit that this ontological interpretation of the completion thesis is unwarranted and ought to be rejected. If we do this, we can and should also reject claim (12) in my initial overview of Sartre's philosophy of action. The resulting position, comprising Sartre's claims (1) to (11), is still substantial and distinctive, and it captures much of what he has to say on the subject. It is not in all respects Sartre's position, but it shares enough with it to warrant the label 'quasi-Sartrean'.

2 Reasons, values, emotions

What can be said in favour of the quasi-Sartrean view? Since I take (1) and (2) to be widely shared, although not entirely uncontroversial, let me begin by examining theses (3) and (4). This will allow me to introduce some of the basic concepts in Sartre's account of reasons-responsive action (henceforth: RA.) By his lights, RA involves four structural elements: (i) an appropriate 'motive' (*motif*), (ii) a choice or decision, (iii) a valued end (*fin*), and (iv) an incentive presentifying that end to the agent (*mobile*). RA being essentially conscious (thesis 2), Sartre takes each of these elements to be registered in consciousness at the time of action, though the agent need not – indeed cannot – have an explicit, conceptually structured, awareness of all of them at the time. In Sartre's terminology, much of the relevant phenomenal consciousness is 'non-thetic' ('non-positional', 'non-thematic').

---

[4] An (intentional) object in the broad phenomenological sense is an item that a subject is conscious of as transcending in its nature or being any one experience of it. Such objects include real, fictional or imagined particulars (including events), types, properties and their instances, states of affairs, aspectual modes of presentation attended to as such, linguistic senses when thought about rather than simply understood, and experiential attitudes (such as imagining, consciously remembering, desiring, being attracted to, etc.) as thought about rather than simply 'lived though', i.e. experienced. Sartre claims, like Husserl, that experiential attitudes, when directly experienced (lived through) rather than thought about are necessarily not given as intentional objects to the subject directly experiencing them. For detailed defences of this claim, see Zahavi (1999) and Poellner (2003).

(i) Turning to the first constitutive element of RA, the *motive* (*motif*), in Sartre's slightly confusing terminology, is the "present state of affairs as it reveals itself to [the] consciousness" of the agent (Sartre 2003: 469). There are two different things Sartre has in mind here. First, the agent needs to grasp the present state of things that is to be changed or actively maintained by his action in *evaluative terms*. That state of affairs with its putative value properties is the *motif*. For example, in the eyes of emperor Constantine the "plebs and the aristocracy of [his] time are corrupt" (Sartre 2003: 469) and this presumed fact is one of his *motif*s.

Second, RAs (except basic ones) involve an understanding of aspects of the situation that are instrumentally relevant, or apparently relevant, to the performance of the action. Introducing a distinction Sartre does not make explicitly, we may call these aspects instrumental *motifs*. An instrumental *motif* is a presumed objective feature or fact making the action good or suitable for the purpose of achieving an end the agent is committed to (Sartre 2003: 469–70). For example, the Merovingian warlord Clovis's instrumental *motif,* his instrumental reason, for converting to Catholicism is that *the church is powerful and will support a king who can help it in its fight against Arianism* (Sartre 2003: 468). RA generally involves a practical, not necessarily thematic, understanding of such *means-end relations*. While Sartre officially defines an agent's grasp of an instrumental *motif* as a presumptively rational justifying consideration (Sartre 2003: 468), and while this fits deliberate and reflective projects like Clovis's quite well, he often acknowledges that the way in which instrumental suitability is grasped in non-reflective action itself is typically not of this propositional kind but rather involves a conscious perceptual registering, in action, of the environment as affording specific opportunities or obstacles. To the voyeur at the door, the keyhole is presented as "to be looked through close by and a little to one side" (Sartre 2003: 283), to the soldier in panic the wall he is running towards is given as *affording shelter* and *to-be-hidden-behind*, the barbed wire in his way as *to-be-jumped-over*, and so forth. It is plausible to hold that in cases where the relevant affordances are highly specific, requiring finely tuned behavioural adjustments, some of these instrumental *motifs* are too fine-grained to be grasped conceptually by the agent;[5] his awareness of them is in

[5] Cussins (2003: esp. 149–59).

these cases non-thetic, even if they are, as they need not always be, attended to.

(ii) A *choice,* or better, a *choosing* is the agent's consciousness of his effective decision or commitment towards attaining an end, and of initiating or maintaining relevant steps towards this end: a normally non-thematic consciousness of effective desire. The choosing is experienced as the agent's *own*, rather than as an alien force, precisely insofar as it is the definitive embracing of an end as *his* end, to be realized or pursued *now* in light of his grasp of relevant *motifs* (Sartre 2003: 471). Note that this characterization does not imply that choosings are necessarily *reflective* or *deliberative*. Many choices in RAs involve neither deliberation nor reflection: agents neither deliberate about possible alternative ends or means and their respective merits, nor does she or her own mental states figure in the explicit intentional content of the choosing. Whatever the correct full characterization of a choosing may turn out to be, for Sartre's purposes it suffices to say that it is that element in an intentional action that distinguishes the latter for the agent at the time, on the one hand, from behaviourally type-identical happenings that are merely passive behaviours (e.g. being pushed) or reflexes, and, on the other hand, from other attitudes towards the same content such as wishes or mere entertainings. It is the aspect of the experience of action that grounds the agent's normal non-observational and non-inferential ability to say, with respect to many of her behaviours, under some descriptions: 'I *did* it'; in contrast to other, possibly type-identical, behaviours that are truthfully reported by her as 'it *happened* to me'; and also in contrast to type-identical possible actions her attitude to which she would report as 'I considered doing it, but didn't actually do it'.

(iii) The third essential aspect of a RA is the action's *end*. Sartre distinguishes between instrumental and intrinsic ends. An instrumental end is what an action aims to achieve in order to achieve a more ulterior end, the latter being what is aimed at for its own sake. Clovis's instrumental end in converting to Catholicism is securing the support of the powerful episcopate in his pursuit of the intrinsic end of ruling over the whole of Gaul. Not all RAs involve distinct instrumental ends, but all have an intrinsic end. The intrinsic end need not be a result of the action but might be some feature of the action itself: an action may be done for its own sake. If the intrinsic end is that for the sake of which the agent performs an action, she needs to have a conscious grasp of it, but Sartre claims that this

does not necessarily imply that an agent has an explicit (conceptual) conscious grasp of all the relevant features that make the end worth pursuing for her.[6] Ends typically involve, although they are not wholly constituted by, objective states of affairs: such things as 'France being liberated' or 'Pierre getting well again' or, in Clovis's case, 'my ruling over Gaul'.[7]

In Sartre's picture, the justifying reasons for a RA depend on whatever is to be said in favour of the action's intrinsic end, and the feature that makes an intrinsic end suitable to justify an action is its being good or *valuable* in some respect (claim 3). Therefore, in order to make an RA fully intelligible I need to *specify* the value that the agent took to qualify its end. What does Sartre mean by 'a value'? He says: "Values ... are demands" (Sartre 2003: 62). "By their nature they 'ought to be' " (Sartre 1999: 88). A positive value is a feature or property that is necessarily such as to merit, pro tanto, being instantiated: it 'ought to be', while a negative value 'ought not to be'. Since, on my reading, Sartre thinks that the question of practical rationality arises and ought to be addressed at the phenomenological level without recourse to metaphysics (thesis 11), his point here should be phrased more perspicuously in a phenomenological idiom: for something to be presented as a positive value is for it to be presented as a property whose nature includes meriting-to-be-instantiated. And the question of whether some possible end E is *correctly* presented as having this property can be answered, by his lights, without incurring metaphysical commitments.

The 'demands' constitutive of value often entail normative demands ('exigencies', Sartre 2003: 60–2) upon agents conscious of themselves as relevantly situated. A *normative* demand is what I am conscious of when I am conscious of an actual or possible *course of action*, as for example, impermissible, acceptable, categorically required etc. What I call a *valuational* demand is acknowledged by acknowledging

[6] According to Sartre, the consciousness of ends is always at least partly non-thetic (Sartre 2003: 115, 485), that is, they are always represented at least in part non-conceptually. For an account of how non-conceptual contents might contribute to a subject's reasons, see Poellner (2003).

[7] Sartre thinks that human ends are never adequately specified in terms of a state of affairs that already obtains: they always involve the envisaging of a condition that is as yet absent or not wholly realized, and in this sense they always aim at a 'not-being' – at something that the agent takes to be not actual at the time of his choosing to act in pursuit of the end. I shall say something about why he thinks this in Section 3.

that it is good (bad) that some possible item should (not) be instantiated, that it merits or does not merit being real – independently of whether it is possible to act in any way relevant to its actualization. For Sartre, normative demands asymmetrically depend on valuational demands (Sartre 1992: 555). Scheler, whose position Sartre adopts on this issue (Sartre 1999: 88), explicitly argues for the priority of the concept of value over the concept of a norm and other deontic concepts, maintaining that we can make sense of a value being actual, being non-dispositionally present here and now, without there being any actual normative demands or normative reasons that could plausibly be associated with it for any agent. It is, for example, not simply unintelligible, not ruled out by the conceptual requirements of evaluative discourse, to take as a non-instrumental value the nature of the 'starry heavens above me', or the basic structure of the physical universe (Scheler 1973: 173–4, 203f). And I can intelligibly acknowledge this as an *actual* value, instantiated now, without being committed to the existence of some non-finite agency that acknowledges or upholds some normative demand in relation to it, or without committing myself to the existence of norms as Platonic, abstract objects. I shall return to Sartre's commitment to the priority of the concept of value in a moment.

(iv) The fourth and final component of an RA is what Sartre calls an appropriate *incentive* (*mobile*). A *mobile* is a conscious 'act' (i.e. intentional mental state) involving an agent's (a) grasping of a possible end as a value, and (b) grasping herself as relevantly situated to pursue that end. A *conclusive incentive* is the agent's understanding of a possible end as the best among the ends presenting themselves to her as relevant in her present situation, which is compatible with the absence of deliberation and with that end being the only end she is conscious of in that situation.[8] Having a conclusive incentive

[8] What about akratic actions in which, on a widespread construal, an agent's effective desires diverge from what the agent takes to be best in the situation? Sartre does not accept that there is akratic action in this sense. The order of valued ends a person is committed to at a time is normally not revealed by his reflective beliefs about his commitments, nor by his reflective second-order desires, but by his choice of first-order ends and, partly dependent upon these, by his affectively perceptual responses to a worldly situation (Sartre 1992: 477–8). The akratic person typically acts against his *beliefs* about his own evaluative commitments, but his affective responses and the actions they motivate disclose the akratic person's real (in *akrasia* typically fragile or self-deceptively disavowed) commitments at the time of action as being different from the content of those reflective beliefs.

> is internally related to *choosing*, i.e. to initiating steps taken in pursuit of the end. An agent's having a conclusive incentive might for instance include (though it would not be exhaustively and fully adequately characterized as) an evaluative episodic belief with the content 'E is supremely beautiful', plus a practical understanding of herself as suitably placed to pursue E *now*. So, in Sartre's picture, the very thing that ultimately seems to justify the action from the agent's perspective – the *evaluative content* of her belief – is also the reason that *motivates* her. Her motivation (her *mobile*) cannot be understood without reference to the evaluative content of her belief. Her incentive includes a mental act with that content and it is *that very content* that makes her incentive the motivating state that it is.

Let me make two initial comments on this account: it is clear that Sartre belongs to the anti-psychologistic camp with respect to both justifying and motivating reasons. There is no problem, in his view, with understanding how a believing can be a motivating mental act *if its content is evaluative*, because sincere categorical evaluative believings just are the kinds of things that, in the absence of countervailing considerations or irrationality, entail inclinations to favour and (where relevant) pursue the content of the evaluative belief. What requires special explanation is not how sincere evaluative beliefs can motivate relevant actions, but how they can sometimes fail to do so. Explanations in such cases may include self-deception, or stronger countervailing beliefs, or the agent's taking herself to be incapable or unpropitiously placed, or various kinds of mental pathologies rendering the agent unable to attempt to act on what she sincerely takes to be good reasons.

My second comment concerns worries about giving a fundamental role to values and evaluation in the theory of justifying reasons. These worries come mostly in two forms. Some would say that the basic justifiers are not evaluative but normative or deontic: they are not of the form 'x is good', but, for example, 'any rational agent ought to ϕ'. I have already said something about this issue and will come back to it in Section 3.

The other worry is whether there might not be reasons that are counter-evaluative or counter-normative. David Velleman asserts that "reasons for acting can be perverse as well. That is, an agent's reason for doing something can be that it's a bad thing to do; and so its justificatory force cannot depend on that of a favourable evaluation."[9] He cites two classics

[9] Velleman (2000: 121).

as illustrations, Milton's Satan and the person in self-destructive despair. Nietzsche's active nihilist might also come to mind in this context, as might a figure like Jean Genet who, according to Sartre's extended existential psychoanalysis of him, chooses to be 'evil' (Sartre 1963: 49–58).

Sartre would rightly insist that it cannot be a justifying reason, a consideration in favour of an action, that there is nothing at all on account of which it merits doing. Examples supposedly illustrating such ostensible reasons are invariably under-described. The individual in genuine self-destructive despair sees himself as irremediably cut off from anything good, as hopelessly immersed in what seems disvaluable to him, and he seeks to escape this condition in the only way that seems left to him. The active nihilist in Nietzsche's sense is typically a kind of radical Manichean: he believes that everything that actually exists is by natural or metaphysical necessity inhospitable to what *would have value*, hence everything real deserves being annihilated. Satan in Milton's version, on some readings, simply desires the apparent good of power; on other readings he is a proto-romantic rebel against a deity that itself has characteristics of a tyrant. The intelligible good vainly aspired to by Genet's choice to be morally evil, according to Sartre, is *being a self-founding consciousness that has an essence*, i.e. being an ontologically self-sufficient substance, in one traditional sense of this term (Sartre 1963: 59–72).

I have said that Sartre does not accept the Humean view that beliefs cannot motivate without the addition of some further independent psychological element that is not itself a belief. I think he is right to reject that view. But there is another cluster of familiar worries about the kind of position he holds. He claims that beliefs can both motivate and potentially justify actions only if they have or imply evaluative contents. The most fundamental of these contents are presentations of potential ends as valuable in some respect. And many of the relevant evaluative contents include actual and potential objective states of affairs in the world essentially involving particulars (claim 4). This seems phenomenologically correct. The valued ends of people's actions are often such things as one's friend being happy, or one's city becoming more attractive, or there being more just government in the world. But if we say that such things can correctly be judged to be valuable, this generates two puzzles, one metaphysical and one epistemological. Metaphysically, it seems to commit us to the idea that the objective world can contain first-order properties, properties had by or involving particulars, that inherently make demands, and these seem strange sorts of properties. Sartre bypasses this metaphysical issue by resolutely sticking to the

phenomenological level. To say that the world as experienced, empirical reality, is correctly presented as potentially containing such properties, and that grasping them can justify actions, does not commit us to a view about the ultimate (metaphysical) status of these properties – no more than does the claim that empirical worldly objects can be correctly presented as having determinate phenomenal colours, and that our perceptual experiences can justify beliefs such as 'this object is scarlet, just as it perceptually seems to me'.

But what about the epistemological problem? How do I grasp a first-order property that inherently makes a demand, that merits or calls for something, *as such*? It would be implausible to say that we access such properties *inferentially* in the basic cases of everyday ends. Not only is it mysterious what the relevant inferences would look like, many of the features that we take to make our ends worthwhile are pretty clearly not grasped by us inferentially. *The badness of a friend's perceived suffering* that makes trying to alleviate it a worthwhile end usually isn't something I access inferentially, nor is *the beauty of this musical performance* that makes getting a recording of it worthwhile.

Sartre's response to the epistemological problem of value properties is a distinctive account of affective intentionality inspired by Scheler (claims 5 and 6). He construes many conscious emotions as essentially involving perceptual acts, and what is perceptually presented in these emotions are value properties of objects, events or worldly states of affairs: "we see that ... these notorious 'subjective' reactions, hatred, love, fear, sympathy ... are merely ways of discovering the world. It is things themselves which suddenly reveal themselves to us as hateful, likeable, terrifying, lovable" (Sartre 2002a: 383–4, also 2004a: 68–9). In *Being and Nothingness* he says, "my indignation has given to me the negative value 'baseness', my admiration has given the positive value 'grandeur'.... Values are [thus] sown on my path as thousands of little real demands" (Sartre 2003: 62).[10] If values are potentially grasped in

[10] Evidently not all conscious emotions are 'ways of discovering the world'. In *Sketch for a Theory of the Emotions*, Sartre at length analyses various emotions as pre-reflectively purposeful attempts to misrepresent the world in situations of difficulty, analogous to the sour-grapes response. But even in that text he recognizes another "main type" of emotion, which is not purposefully distorting in this way, but whose evaluative content is "motivated by the object itself" (Sartre 2002b: 57). Richmond (2011: 153–5) rightly observes that this second type of emotional experience plays a much more central role in Sartre's thought than is acknowledged in the *Sketch*, not only – I would add – in his account of being-for-others but in the foundations of his phenomenology of value.

conscious intentional emotions whose contents they are, then it is clear how Sartre can meet the internalist requirement that on his analysis is built into the concept of value; for such emotions are intrinsically motivating states. They are incentives (*mobiles*) and they are motivating, on his view, because their evaluative contents are not merely believed but also intuitively presented in them. Sartre therefore has an explanation of why an emotional presentation of a value or disvalue is inherently motivating, although this motivation often may be no more than an inclination, while a mere belief with the same evaluative content need not be. In emotional presentation, the relevant value appears to be intuitively present, either directly (perceptually) or indirectly (imaginatively), while this is not the case in what phenomenologists call empty (i.e. merely symbolically mediated) conscious belief. A severely depressed person may not be to any degree motivated to act on the basis of his sincerely held evaluative and normative beliefs, even believing that he is well-placed to act, precisely because he cannot 'feel' these values, however important he may believe them to be.

But can affectively presented contents also be potentially rationalizing or justifying? The cognitivist view that many emotional episodes have contents, that their contents are typically evaluative, and that they are individuated in terms of their evaluative contents, has gained much support in recent decades. But what even many of its supporters balk at is the idea that some of the relevant evaluative contents, in the basic cases, might be *perceptual*.

Consider one of Sartre's own examples: a person who takes flight in fear at a lethal threat, perhaps a soldier fleeing from an enemy assault (Sartre 2003: 465). Sartre wants to say that the soldier correctly apprehends *through* his fear a certain event (the enemy attack) as instantiating a disvalue, as something that ought-not-to-be, and this affective perception in his case presents avoidance of that threat, making it disappear by running away, as something he should do. The fear, insofar as it is a conclusive incentive for this person, also reveals and depends on his dominant intrinsic end at the time: preserving his life. It reveals the preservation of his own life as an overriding value for him (Sartre 2003: 459, 465). This end is typically not thematically manifested in his fear, but is implicit in his terrified apprehension of the enemy assault as lethally threatening to him and as a conclusive reason for taking flight. Being implicit, the valued end is *available* for explication: if the soldier is subsequently asked about his end at the time, and if he is not in bad faith and does not suffer from memory failure or such like, he will concede that the fleeing was something he did, and that his end

was indeed the preservation of his life. The soldier's behaviour is therefore a reasons-sensitive action, involving a choice in the light of non-reflectively apprehended reasons. Some of these reasons are provided by the contents of affective perceptions of disvalues in the world; but these conscious perceptual contents are themselves partly determined by his non-reflective 'projective' consciousness of a valued end to be attained in the future.[11] For Sartre, then, value, practical normativity and choice are at this basic level non-reflective phenomena – he plausibly holds that they are not *generated* by reflective distancing.

One question raised by this account concerns the way in which the valued end – in the example, the preservation of the soldier's life – is given to the agent. If one is sympathetic to the idea that aiming for a certain end is, in a case like this, not simply a contingently related effect of the conscious emotion of fear but part of what constitutes that affective state, one incurs an obligation to specify how that end is present to the agent. Sartre says that it is implicitly given ('non-thetic'), but that by itself is not very illuminating. He does not want to say that the given-ness of the end is *identical* with the affective perception of the enemy assault as irresistibly and terrifyingly dangerous, but holds that the specific character of that perception *depends on* the agent's commitment to this end (Sartre 2003: 457, 465, 477–9.). The latter point at least seems correct. Presumably, someone who is largely indifferent to his life, or has other, firm overriding commitments, will not affectively perceive the event in this way, but experience it (say) as excitingly or somewhat unpleasantly dangerous; his affective response will at any rate not be such unqualified terror. Now, one might be tempted to regard the agent's representation of his end as an *unconscious* standing commitment and that unconscious mental state as at least partly responsible for the specific terrifying way in which the event registers in his affective consciousness. Sartre, however, insists that the valued ends fundamentally motivating and justifying reasons-sensitive actions are not unconscious (analogous to blindsight) but implicitly conscious. It seems that there are only two theoretical options compatible with this claim. One might either say that the implicitness of the end simply consists in the fearful event being experienced as utterly terrifying and as imposing

[11] The fact that the content of a putative affective perception partly depends upon an agent's ends need not impugn its possible veridicality, because this is true also of ordinary sensory perception: a person's being distracted by a certain project may prevent him from noticing features in his environment that are nevertheless there to be perceived; his adopting different ends might conduce to making him attentive to those features.

a categorical normative demand: 'you must run'. But this option falls foul of Sartre's oft-repeated tenet that no such ('free') action is merely imposed or demanded by *present* features of the object-world, and that its motivation depends, rather, upon an awareness of some as-yet-absent valued end (Sartre 2003: 457–8). The remaining option would be to say that an emotional episode like the soldier's fear includes not only a perception of a present disvalue in the world, but also an 'implicit' awareness, distinct from this, of the valued end, and that this latter aspect affects the character of the former. Sartre's subtle but often neglected theory of imaginary intuitive presentifications of what is absent via 'analoga' might help towards explaining this kind of consciousness of value (Sartre 2004a: 68–83). I cannot attempt this task here and must leave this issue unresolved for now, not without noting that addressing it seems to me one of the most important tasks of a phenomenological theory of action.

Sartre's theory of affective value perception does of course not entail the view that *all* conscious intentional emotions are 'world revealing' or that they all disclose genuine reasons. Just as there are illusory or hallucinatory pseudo-sense-perceptions, so there are illusory or delusional conscious emotions: phobias, compulsions, educationally ingrained narcissistic affects, and so forth. But in an example like that of the soldier, we would be hard put to deny that the emotion makes intuitively accessible genuine pro tanto reasons. The question of whether he is right to take them as *conclusive* reasons, and what would be required for him to recognize them as not conclusive, raises a host of issues, some of which I shall come to presently. For now, let me summarize the requirements which a Sartrean theory of some emotions as, in part, perceptual experiences of values would have to satisfy. Emotion episodes can count as including value perceptions just in case they meet the necessary and jointly sufficient conditions of perceptual experiences more generally. These are, I submit, as follows:[12]

(i) perceptual experiences present apparent phenomenal properties of objects (in the case of affective perceptions, pre-eminently value properties);
(ii) they are non-doxastic (i.e. not judgements or dispositions to judge) but make contents available that can be captured in appropriate judgements;

[12] I here use 'perceptual experience' in a broad sense, including partial illusions.

(iii) they are epistemically direct;
(iv) if they are experiences as of particulars, they systematically appear caused by property-instances presented 'in' or to them; and
(v) their condition of success is veridicality.

I shall not discuss here whether these conditions can be met by some emotions, having done so elsewhere.[13] What I want to turn to instead is the cluster of claims (7) to (10) from the initial overview of Sartre's position. These propositions delineate something like a hierarchy of values in which the value of authentic consciousness occupies a pre-eminent place.

3 The value of authentic consciousness

Consider theses (7) and (8) from the initial outline of Sartre's theory:

(7) The justifying force of the values of objects is trumped by the value of reasons-responsive consciousness as such.
(8) Insofar as consciousness is adequately reasons-sensitive, it is fundamentally guided by normative demands deriving from the value of such consciousness itself.

By 'consciousness' in the strict sense relevant here, Sartre means lived *experiential attitudes* or modes of consciously relating towards various contents: conscious believings, doubtings, desirings, imaginings, affective attitudes and so forth, as these are given in the first-personal 'lived' perspective. Consciousness in this sense is 'present to' itself – we are conscious not only of intentional contents but also of our experiential attitudes towards them – but it is, in its fundamental mode of self-presence, not given to itself as an intentional object (Sartre 2003: 6–12). This is in part what Sartre means when he says that consciousness is 'beyond [object-]being' or 'beyond the world'. In its most concise, albeit elliptical, form, Sartre's argument for the overriding value of

[13] Poellner (2007). Other authors sympathetic to the affective perception model give somewhat different lists of the conditions an emotion would have to satisfy to count as perceptual; see Goldie (2007) and Tappolet (2012). I think that these are not sufficiently demanding, making their accounts vulnerable to the objection that a mental state might satisfy the proposed conditions without being genuinely perceptual.

reasons-responsive consciousness is set out in the following passage in *Existentialism and Humanism*:

> I declare that freedom, in respect of concrete circumstances, can have no other end and aim but itself; ... when once a man has seen that values depend on himself, ... he can will only one thing, and that is freedom as the foundation of all values ... Obviously, freedom as the definition of a human being does not depend upon others, but as soon as there is commitment, I am obliged to will the freedom of others at the same time as mine. I cannot make freedom my aim unless I make that of others equally my aim. (Sartre 1980: 51–2; cf. 1992: 414)

This is a popular lecture, and Sartre's formulations in it sometimes sacrifice precision for rhetorical effect. Some modifications and clarifications are needed to extract his actual argument. First, he is not entitled to talk about reasons-sensitive consciousness as 'freedom' in his premises if the argument's *conclusion* is to be that such consciousness is free or autonomous. Second, while there is, by his lights, a conceptual and existential dependence of actual value on consciousness – for only consciousness can register the 'demands' that are constitutive of values and there can be no actual demand not registered by anything or anyone (Sartre 2003: 62) – there is, according to the dominant strand in his thinking, no dependence of value on *any one* individual consciousness. Let me, then, offer the following as a reconstruction of what Sartre should have said, consistently with his considered overall position:

(i) The more important values I rightly recognize depend existentially on the reasons-sensitive consciousness (henceforth: RC) of myself and others.
(ii) Insofar as I acknowledge such values, I cannot consistently not 'will' – i.e. affirm, endorse as an end – RC in myself and others.
(iii) I cannot rationally not acknowledge any such values.
(iv) Hence I cannot consistently not will RC universally, i.e. in each instance.
(v) I must take RC, in each of its instances, as the primary value for the sake of which I act – as my primary end.

Elsewhere, he adds:

(vi) The value of self-transparent and adequately reasons-acknowledging (i.e. *authentic*) consciousness is 'absolute': it is greater than

> the aggregated value of anything that is not such a consciousness. (Sartre 1999: 96)

And:

(vii) For all I know, every being characterized by RC is capable of authenticity. (Sartre 2003: 475–86)

If one accepts this argument so far, it is tempting to conclude:

(viii) If I am adequately rational, I regard every being characterized by RC as having absolute value.

If this argument goes through, we can see why an advocate of the quasi-Sartrean view sketched here should feel entitled to claim that an adequately rational consciousness is 'free' in the sense of *autonomous* or *self-determining*: such a consciousness would be fundamentally motivated by its grasp of the overriding value of adequately reasons-sensitive consciousness as such, and it seems a perfectly intelligible and not wildly revisionary application of the concept of self-determination to say that just in case a consciousness is motivated in this way, it is determined by itself.

One issue about Sartre's argument as I have reconstructed it is the move from (ii) to (iv): the unstated supplementary premise here is that the value of RC accrues to it on account of its 'foundational', i.e. constitutive, role, and this role is, as Sartre puts it, a 'universal': if I value something on account of its having that role, I am rationally committed to valuing any instantiation of that role. I won't question this part of the argument. In the remainder of this chapter, I instead want to focus on the move from (iv) to (v) and then to (vi). Both moves are clearly invalid without additional premises. What might these premises be?

Given the apparent affinity of Sartre's conclusions with Kant's, it is tempting to look for them in the Kantian tradition. But it is at least doubtful whether the sort of considerations that we find there should appeal or are even available to Sartre. Let me mention three of these:

(1) It is sometimes said by Kantians that the only intrinsic (here: non-relational) value is that of rationally choosing consciousness itself.[14] If this is right, it follows straightforwardly that nothing else

[14] Korsgaard (1996: esp. 256–62).

can compete in intrinsic value with rational consciousness. Sartre cannot say this because intentional consciousness is on his analysis itself relational: if there is such consciousness, then, necessarily, it is embodied and situated in a world of value-laden objects. And neither these objects nor their apparent phenomenal properties, including their values, are with phenomenological plausibility construed as immanent to any one or all consciousnesses aware of them. The relation of intentional consciousness and worldly significance is one of necessary *correlation* and mutual dependence.

(2) Another claim that is sometimes made by Kantians is that the authority of any values other than rational consciousness is a derivative or borrowed authority. Only insofar as I acknowledge the authority of rational consciousness can such other values be presented to me as authoritative.[15] This point might then be given an ontological, not merely an epistemological, reading to the effect that any other values are *constitutively* dependent on the value of rational consciousness. But this ontological interpretation is not obviously rationally mandatory. Why should we not be entitled to say that the value of rational consciousness is, at least in some cases, epistemological and instrumental, allowing one kind of access to non-instrumental values *some* of which might, in principle, be accessible also to a non-rational consciousness? At the least an additional argument would be needed to rule out this apparent possibility.

(3) Here is one such supplementary argument: it might be said that reasons-responsiveness is a constitutive feature of intentional agency, and that I cannot intentionally dissociate myself from my intentional agency without practical inconsistency, since such agency *constitutes me* and I cannot dissociate myself from myself.[16] In this sense, my commitment to RA has to be an unconditional commitment to it as a non-instrumental value. Sartre might seem to be making a similar claim when asserting that I cannot consistently reject my freedom (Sartre 1980: 51). Application of the universalizing move underpinning proposition (iv) in his reconstructed argument above might then be taken to yield a rational requirement to value reasons-acknowledging consciousness *unconditionally* wherever it is instantiated.

---

[15] Korsgaard (2009: e.g. 23–5, 116).

[16] Cf. Korsgaard (2009: 1–2, 180).

But it does not follow from this that adequately reasons-responsive consciousness has to be acknowledged by me as having pre-eminent value, let alone *absolute* value in Sartre's sense, if I am rational. There is nothing inconsistent in thinking that rational thought might yield the conclusion that there are other values that trump it. Presumably it is this very idea that is shared by historically widespread and diverse forms of monistic mysticism such as Vedanta and Buddhism and its Western counterparts in Schopenhauer and (perhaps) Nietzschean Dionysianism. They all hold that I can rationally renounce my rational agency and thus, in a sense, renounce myself for the sake of values that are 'higher' than such agency or selfhood. These ideas may be wrong, and indeed Sartre himself would not accept them since he (mistakenly) takes the notion of a non-projective, not end-directed consciousness to be incoherent, but they are not false on purely a priori grounds generated by practical commitments necessarily incurred by us merely qua formally rational agents.

I suggest that Sartre would concur with this last point, and this may be a further motivation of his claim that evaluative concepts are more basic than normative or deontic concepts: we can grasp the concept of value independently of normative concepts such as the concept of a *practical reason*, but not vice versa. There might be goods in a world without rational beings. Or, to translate this point into the metaphysically non-committal phenomenological register which, I have argued, captures everything that is practically important about Sartre's approach: something might be presented as valuable and as actually so, without anything being actually presented *as a reason*. Presumably this is what is meant to happen, for example, in the conditions of *moksha* and *samadhi* in some versions of Eastern mysticism, or in the Dionysian states envisaged by early Nietzsche or (possibly) in Wagner's *Tristan*.

Sartre, and the quasi-Sartrean view, are therefore also committed to rejecting currently popular 'buck-passing' accounts of value, such as the idea that *being valuable* is the higher order property, possessed by some non-evaluative ('natural') properties, of *providing reasons* to respond in various positive ways towards those lower-order properties.[17] According to the quasi-Sartrean view, the basic practical reason-providers cannot be adequately described without the use of evaluative terms. They are not 'natural' (non-evaluative) properties but rather evaluative characteristics, that is, characteristics that are adequately presented to an

[17] Scanlon (1999: e.g. 95–7).

agent only if they are presented *as* meriting (pro tanto) being favoured or disfavoured.[18]

The fundamental issue between a value-centred view, of which Sartre's is one version, and reasons-centred views such as Kantian or buck-passing accounts, would seem to be whether anything can in principle be given to a consciousness *as meriting* a certain response without being presented to it as providing a reason. It is not clear why this should be a priori impossible, at least if reasons are taken to be fully conceptually structured items such as propositions or facts. An advocate of the value-centred view may hold that a consciousness whose phenomenal 'contents' are not fully or determinately conceptualized might still be aware of these incompletely conceptualized contents as meriting a certain response. Perhaps being happy with or acquiescing in being attracted by such a conceptually indeterminate content can be thought of as an example of such an awareness. This seems, at any rate, to be a typical feature of the sort of self-forgetful, 'mystical', or 'Dionysian' states mentioned above as subsequently described by their adepts.

If none of the Kantian ideas referred to earlier are suitable to fill the gap in Sartre's move from (iv) through (v) to (vi), can it be filled? It seems that the unstated further assumptions supposedly legitimating this move are based on his thoughts about an essential feature of intentional consciousness that was briefly broached in Section 1: intentional consciousness aims at *completion*, understood as the experiential *absence of lack* (Sartre 2003: 111–15). The end-directed, projective character of agency implies that, qua agents, we are always aware of the present as lacking or deficient in some respect. Any conscious pursuit of a project necessarily entails such an (often unthematic) awareness of an undesirable lack, even if the lack in question is only the fragility of a good currently present, its needing to be actively safeguarded against possible loss or destruction. The for-itself's projective structure is therefore

[18] Sartre can agree with the buck-passers that the correct description of X as 'being good' does not provide additional reasons beyond the reasons supplied by other properties of X. But this is because the reasons-providing properties are (not non-evaluative but) more determinate evaluative properties, such as something's *being beautiful*, or *being just*, or *being invidious*. Note that a property often cited as non-evaluative by contemporary philosophers, the property of *being (sensorily and phenomenally) painful*, is an evaluative property by Sartre's lights, and he seems right on this. I cannot be directly aware of a sensation as unqualifiedly painful without being aware of it as pro tanto meriting to cease, and this is not a contingent psychological fact (Sartre 2003: 357, 408–9). If anyone is inclined to dispute this, they should try genuinely to envisage a sensation that they would regard as strongly and unqualifiedly painful. For most people, the sensation produced by a dentist's drill in a healthy tooth without anaesthetic should be a suitably vivid case of this sort.

tantamount to the ever-renewed endeavour to overcome or remove an undesirable conscious lack or 'incompleteness'. Sartre sometimes calls the desired experiential state of completion 'substantiality'. In the *War Diaries* he says about this: "the source of all value, and the supreme value, is ... substantiality" (Sartre 1999: 111; 2003: 112–13). This idea is by no means only a motivated error characteristic of the for-itself in bad faith. Authenticity "is not a question of [consciousness's] seeking any value other than substantiality – if it did, it would cease to be *human* consciousness" (Sartre 1999: 112; cf. 2003: 115).

That substantiality is the 'supreme value' appears to be regarded as self-evident by Sartre, which suggests that he takes it to be entailed by the analytic truth that the highest conceivable value is the universal absence of a deficiency in value, in conjunction with his claim that all actual value is experienced value. (I shall call the latter claim the experience-dependence thesis about value, henceforth EDV). But if that is Sartre's thinking, it requires a distinction, which he himself does not explicitly draw, between the concepts of substantiality and of veridical substantiality. The former is the concept of a consciousness that experiences itself as without lack and, in this sense, as completed. But clearly the presence of such an experiential state – if it were possible – would not be equivalent to the absence of value-deficiency, even granting EDV, for the content of such a state might be mistaken. I might experience myself and the world as lacking nothing, but this experience may evidently be illusory. So, given EDV, what is self-evidently of supreme value is not substantiality per se but veridical substantiality: the absence of a deficiency in value, registered as such in each individual consciousness. (If it were not registered thus in some consciousness, that consciousness would necessarily experience lack and thus be 'non-substantial', hence the world would be, to that extent, deficient in value.) The distinction between substantiality *simpliciter* and veridical substantiality enables us to make sense of Sartre's claim that in authenticity, the commitment to the value of substantiality is not abandoned but 'corrected' and purified' (Sartre 1999: 112). Authentic (self-transparent) consciousness, like inauthentic consciousness, is constitutively oriented towards, motivated by, the value of substantiality or completeness, but unlike inauthentic, inadequately rational consciousness, it acknowledges (a) that substantiality is an 'unrealizable' for finite consciousness,[19] and

[19] The unrealizability of consciousness's completion is the source of one of the deepest tensions in Sartre's philosophy. As I noted in Section 1, he sometimes says that such completion (substantiality) is a *self-contradictory* end. If he means

it acknowledges (b) the universality of value: it acknowledges that, if something is a (dis)value, it is a (dis)value wherever it is instantiated.

It seems that Sartre's thinking underpinning his claim that fully rational and authentic consciousness has 'absolute' value is this: such consciousness is oriented in the right way towards an unqualifiedly valuable end: a veridical substantiality, and hence the completion not only of one's own consciousness but also of that of others. But this end is contingently unrealizable. The best that is attainable in the real world is therefore consciousness's *orientation towards* this end: "Subjectivity finds its meaning ... in this Good, which never is and which is perpetually to-be-realized" (Sartre 1992: 556). "Thus the source of all value, and the supreme value, is ... substantiality ... [it] forms part of human nature, but only in the capacity of a project" (Sartre 1999: 111).

If this is indeed Sartre's thinking, it is not obviously plausible without other assumptions. Consider: if immortality were the highest value but were in fact impossible to achieve in the actual world, it would not follow that the best thing in the actual world would be the *striving* for immortality. The best thing in that world would, presumably, be extreme longevity. What Sartre needs but fails to give is a more contentful *phenomenological* account of what consciousness's 'completion' might consist in which, if he is right, would show there to be an essential connection between the supreme value of veridical substantiality and the value of being motivated by the pursuit of it.

I shall conclude with some indications of what such a phenomenological account would need to look like, consistent with much of what Sartre himself says except for his problematic ontological interpretation of completion (as being an in-itself-for-itself). A consciousness that has attained completion would have to be a consciousness that is not aiming to achieve a yet-absent end, because all its ends would have been attained; it would take everything really desirable to have been actualized, which entails that what it takes to be most valuable would be given to it as realized. Now, Sartre has asserted that an adequately reasons-sensitive

---

by this that it is incompatible with the very idea of phenomenal consciousness, then it clearly cannot be a source of justifying reasons at all. A weaker, but philosophically more attractive, claim would be that an experience of completion or substantiality could not be a (wholly) *intentional* mode of experience and that it is in fact humanly unattainable. A completion of consciousness would involve a fundamental transformation of it, but it is not a priori incompatible with phenomenal consciousness ('what-it-is-likeness') as such. This way of taking what Sartre should have said fits best with the passages from the *War Diaries* cited above.

consciousness cannot take anything other than authentic consciousness itself to be most valuable, hence, *ex hypothesi*, the attained end of a 'completed' consciousness could not be given to it as (exhaustively characterizable as) an intentional object, since consciousness as lived is necessarily not an intentional object. Further, since what is most valuable is not just *one particular* for-itself's combined authenticity (Sartre 1999: 96) and completion (Sartre 1999: 110–12; 2003: 112–13), but such consciousness wherever it might be instantiated – i.e. the totality of consciousnesses, converted to authenticity and completed – a veridically substantial authentic consciousness, if it were possible, would have to be one that would be aware of *all* consciousnesses as authentic and completed. Moreover, it would have to be aware of the value of all this intuitively, not merely via some symbolic (e.g. linguistic) representation, for such representations on their own essentially presentify some *intentional object* as *absent*, as not experientially present to the subject. Since Sartre holds that all intuitive awareness of something as a value is essentially affective, consciousness's completion would have to be a kind of affective awareness, one that could appropriately be described as an affective experience as of 'perfection'. In fact, it seems that the consciousness in question would have to be rather like the experiential life that traditional theology has attributed to the post-mortal souls of the beatified, and it is therefore after all not wholly inappropriate for Sartre to characterize it as a divinized consciousness (Sartre 2003: 114), albeit not for the reasons given by him. One may well think that a *veridical* consciousness of this sort is metaphysically impossible in this actual world, although, *pace* Sartre, it is not logically impossible.

If this description of what a veridically 'substantial' consciousness would have to be is phenomenologically along the right lines, what can be said on behalf of Sartre's claim (a) that this "supreme value ... forms part of human nature, but only in the capacity of a project" (Sartre 1999: 111)? It is this claim, I have suggested, that seems to motivate his assertion (b) that each instance of authentic consciousness that pursues that project in the right way instantiates a special kind of value greater than any object-value. It seems clear that (a) only supports (b) if the pursuit of that project itself instantiates, if imperfectly or to a lesser degree, the very features that constitute the supreme value of veridical substantiality. If it did not do this, why should it partake of or approximate to the pre-eminent intrinsic value of such substantiality rather than having merely instrumental worth relative to it?

An analogy may help to make this point more vivid. Kant, on one reading, argues that supreme value would reside in a holy will, a will

motivated by the content of the categorical imperative without that content needing to be presented to it in imperatival form, since such a will's inclinations would be spontaneously in harmony with it. But an unqualified worth, for Kant, accrues also to a practically rational will that is less than a holy will but is consistently motivated by duty. And, presumably, even an imperfectly rational consciousness intermittently in thrall to non-moral inclination while striving, sometimes successfully, to be motivated by practical reason – even such a consciousness has an incommensurably higher value, in the Kantian view, than anything that is not rational at all. It is hard to see what could justify such a view if not the putative fact that the imperfectly rational consciousness shares an essential feature, albeit only intermittently, with the holy will – to wit, its being motivated by its grasp of the content of the categorical imperative.

An analogous point applies to Sartre's claims. According to the present interpretation, his remarks on the supreme value of (veridical) substantiality in the *War Diaries* underwrite his statement in that text that authentic consciousness has 'absolute' value and supply the unstated premises of his argument in *Existentialism and Humanism* that each instance of it has to figure as a 'primary end' for authentic and adequately reasons-responsive consciousness. If this argument is to succeed, Sartre, or an advocate of the quasi-Sartrean view, would need to show how value features of a condition (veridical substantiality) that is metaphysically or contingently unrealizable can be shared by a for-itself's project to approximate to such a condition, but cannot be shared by anything else. This would require, among other things, a more concrete phenomenological description of the mode of affectivity that would constitute an intuitive awareness of those value features and would thereby constitute 'completion'.

References

Cussins, A. (2003) "Content, Conceptual Content, and Nonconceptual Content", in *Essays on Nonconceptual Content*. Ed. Y. H. Gunther. Cambridge, MA: MIT Press.

Goldie, P. (2007) "Seeing What is the Kind Thing to Do: Perceptions and Emotions in Morality", *Dialectica,* 61: 347–61.

Korsgaard, C. (1996) "Two Distinctions in Goodness", in *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

—— (2009) *Self-Constitution*. Oxford: Oxford University Press.

Poellner, P. (2003) "Non-Conceptual Content, Experience and the Self", *Journal of Consciousness Studies,* 10: 32–57.

—— (2007), "Affect, Value, and Objectivity", in *Nietzsche and Morality*. Ed. B. Leiter and N. Sinhababu. Oxford: Oxford University Press.

—— (forthcoming) "Early Sartre on Freedom and Ethics", *European Journal of Philosophy*.
Richmond, S. (2011) "Magic in Sartre's Early Philosophy", in *Reading Sartre*. Ed. J. Webber. London and New York: Routledge.
Sartre, J.-P. (1963) *Saint Genet: Actor and Martyr*. Tr. B. Frechtman. London: W.H. Allen & Co.
—— (1980) *Existentialism and Humanism*. Tr. P. Mairet. London: Methuen.
—— (1992) *Notebooks for an Ethics*. Tr. D. Pellauer. Chicago: University of Chicago Press.
—— (1999) *War Diaries. Notebooks from a Phoney War 1939–40*. Tr. Q. Hoare. London: Verso.
—— (2002a) "Intentionality: A Fundamental Idea of Husserl's Phenomenology". Tr. J. P. Fell, in *The Phenomenology Reader*. Ed. D. Moran and T. Mooney. London and New York: Routledge.
—— (2002b) *Sketch for a Theory of the Emotions*. Tr. P. Mairet. London and New York: Routledge.
—— (2003) *Being and Nothingness*. Tr. H. Barnes. London and New York: Routledge.
—— (2004a) *The Imaginary*. Tr. J. Webber. London and New York: Routledge.
—— (2004b) *The Transcendence of the Ego*. Tr. A. Brown. London and New York: Routledge.
Scanlon, T. (1999) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
Scheler, M. (1973) *Formalism in Ethics and Non-Formal Ethics of Values*. Tr. M. S. Frings and R. L. Funk. Evanston: Northwestern University Press.
Tappolet, C. (2012) "Emotions, Perceptions, and Emotional Illusions", in *Perceptual Illusions*. Ed. C. Calabi. Basingstoke: Palgrave Macmillan.
Velleman, J. D. (2000) "The Guise of the Good", in *The Possibility of Practical Reason*. Oxford: Oxford University Press.
Zahavi, D. (1999) *Self-Awareness and Alterity*. Evanston: Northwestern University Press.

8

Kantian Radical Evil and Sartrean Bad Faith

Justin Alam

1 Introduction

There is a particular problem in Kant's doctrine of evil, one which has the potential seriously to impact upon both his account of rational agency and his moral philosophy. In this chapter, I would like to offer a solution to this problem. Briefly stated, the issue arises because of the following considerations: Kant believes that our actions are guided by the maxims we have endorsed and that we are ultimately guided in our choice of these maxims by a freely chosen supreme maxim. This can only be good or evil[1] and constitutes a person's overall moral disposition. The difficulty for the doctrine of evil is that as free and rational beings, we have overriding reason to choose the moral supreme maxim because it (and only it) allows us to affirm both our rationality and our full freedom (autonomy). This makes it mysterious how any will could bring itself to choose the opposite: the evil supreme maxim.

The reason this problem in the account of evil is also problematic for Kant's theory of rational agency is that if we are unable to choose the evil supreme maxim, then it obviously cannot play the role for which it was intended in the theory (i.e. providing ultimate guidance). In addition, such a debarment would also be an embarrassment for Kant's moral philosophy because in being left with only the moral supreme maxim, Kant would be saddled with the theoretically undesirable corollary that

[1] Kant believes each of us must, as a matter of fact, either endorse and be guided by a supreme maxim which prescribes the prioritization of duty over self-love or one which prescribes the opposite. In the former case, the person is morally good, and, in the latter, she is evil. He believes it is impossible to fail to possess one maxim or the other and that it is impossible to endorse both at the same time.

no one is evil and that all people have good moral dispositions, and Kant at least seems to believe the opposite: that everyone is evil (R 6: 32).

My solution to the problem of how a will could take evil to be pre-eminently choice-worthy despite knowing that a choice of morality is overriding is that it has *deceived itself* into taking evil as the policy which affirms its own nature as freedom itself. However, this putative solution is itself problematic as self-deception is *prima facie* paradoxical in at least two ways: first, self-deception seems to involve holding two contradictory beliefs. In this case, the first belief would be that an overarching policy of evil is supremely choice-worthy and the second, that morality is. The other paradox is that it at least seems that the self-deceiver must both be aware of the process or act of deception (as deceiver) and ignorant of it (as dupe) at the same time. I argue that Sartrean bad faith – suitably interpreted – dissolves these paradoxes. With this account, an unparadoxical conception of self-deception can be applied to an account of why the Kantian will chooses evil despite the overridingness of morality.

2 Background – Kant's theory of rational agency

It may be useful to outline the interpretation of Kant's model of rational agency with which I am working (the Incorporation Thesis as explicated by Henry Allison). Kant believes that instead of being impelled by our desires – our incentives or *Triebfedern* – we freely choose whether to accept them as reasons for action. This applies whether the agent is motivated by duty to do what is morally required or by inclination to do something which morality does not enjoin (or which it also positively forbids). If an incentive is endorsed, the agent thereby adopts a maxim – a subjective practical principle. The idea that an act of incorporation is required for action is clearly shown in a key quotation Allison (1990: 39–40) provides from *Religion Within the Boundaries of Mere Reason*: there, Kant says, "freedom of the will is of a wholly unique nature in that an incentive can determine the will to an action only insofar as the individual has incorporated it into his maxim (has made it into a general rule in accordance with which he will conduct himself)" (R 6:24).[2]

[2] I cite Kant's works in parentheses in the text, referring to the volume and page number of the Prussian Academy edition of Kant's *Gesammelte Schriften* (1902–), using the standard abbreviations. The texts cited and their abbreviations are as follows: the *Critique of Practical Reason* (KpV), *Groundwork of the Metaphysics of Morals* (G), *Religion Within the Boundaries of Mere Reason* (R), and *The Metaphysics of Morals* (MS). The English translations of these works from which I quote are listed in the References section.

Since the will is precisely the agent's *rational* power of choice, it makes sense that a will requires a *reason* to incorporate an incentive into a maxim (i.e. to adopt a maxim of action). These reasons are provided by higher order maxims, and this is why Kant supposes an agent's maxims are arranged in a hierarchy. The hierarchy is headed by a supreme maxim in which the agent either prioritizes self-love over duty (the evil supreme maxim) or duty over self-love (the moral supreme maxim). In the *Religion*, Kant argues that this supreme maxim must contain both the incentive of duty and "the incentives of self-love and their inclinations" (R 6:36) because we are both moral beings and sensuously affected ones. As a result,

> the difference, whether the human being is good or evil must not lie in the difference between the incentives that he incorporates into his maxim (not the material of the maxim) but in their *subordination* (in the form of the maxim): which of the two he makes the condition of the other. (R 6:36)

In other words, it is not the case that the good have only the incentive of duty and the evil have only the incentives of self-concern in their supreme maxim. Rather, everyone has both of these within this maxim and, therein, the good prioritize the incentive of duty over the concerns of self-love and the evil do the opposite.

We can already see in the second *Critique* the notion of a supreme maxim as the ground of the choice of lesser maxims of action in the following passage in which Kant is making just that claim with regard to the maxim of self-love:

> Now, a rational being's consciousness of the agreeableness of life uninterruptedly accompanying his whole existence is *happiness*, and the principle of making this the supreme determining ground of choice is the principle of self-love. Thus all material principles, which place the determining ground of choice in the pleasure or displeasure to be felt in the reality of some object, are wholly *of the same kind* insofar as they belong without exception to the principle of self-love or one's own happiness. (KpV 5:22)

Kant is more explicit about the idea that a freely chosen supreme maxim must be the ground of lesser maxims in the *Religion*. Having made the point that the subjective ground of the exercise of freedom must itself be freely chosen (or else no action can be imputed), he says:

> Hence the ground of evil cannot lie in any object *determining* the power of choice through inclination, not in any natural impulses, but only in a rule that the power of choice itself produces for the exercise of its freedom, i.e., in a maxim. (R 6:21)

This is closely followed by the point that

> [w]henever we therefore say, "The human being is by nature good," or, "He is by nature evil," this only means that he holds within himself a first ground (to us inscrutable) for the adoption of good or evil (unlawful) maxims. (R 6:21)

Next, according to Kant's strictures, everyone must choose one of the two possible supreme maxims – one must have some ultimate guiding light or other in order to make reasoned choices. However, given that the two available maxims take us in opposite moral directions, it is not possible to choose both (i.e. to be a syncretist); that would be equivalent to being committed to both good and evil at the same time, which is impossible. Finally, it is because the supreme maxim guides agents in their choice of maxims – personal rules which can be fairly fundamental to the way we lead our lives – that it is considered equivalent to a person's moral disposition.

3 Motivations for Morgan's rational reconstruction

There are a number of problems in Kant's account of radical evil as expounded in the *Religion*. First, Kant posits a *universal propensity to evil* in the human will. Now, to claim that this propensity is universal might not be problematic if it were taken to be a mere *susceptibility* to evil. The thought would be that no one is ever entirely above temptation. However, some of Kant's remarks about the propensity to evil suggest that he thinks it is identical to possession of the evil supreme maxim. For example, he says that the propensity to evil is "the formal ground of every deed contrary to the law" (R 6:31). Providing grounds for deeds is the sort of role we would expect of a maxim, and providing such a fundamental ground as Kant describes here is the sort of role we would expect of a supreme maxim. But we have seen that to have the evil supreme maxim is to have an evil *disposition*. So it seems that having a propensity to evil might be the same as possessing the evil supreme maxim, which, in turn, is the same as having an evil disposition. This means that by claiming that the propensity to evil is universal, Kant is

effectively saying everyone has an evil disposition – that we all have evil characters.

Second, Kant supposes that the propensity to evil is inextirpable. Again, this claim is perhaps plausible if taken to mean a mere susceptibility to evil: this would then be the modest claim that no one can ever guarantee that they have rid themselves of the possibility of temptation. But inextirpability of the propensity to evil is less plausible if Kant is taking that propensity to be the same as possession of the evil supreme maxim (again, because he also takes the latter to be constitutive of an evil *disposition*). Given the apparent identification, Kant seems effectively to be claiming that no one can rid themselves of their evil disposition and acquire a good one. However, we might think that bad people can become good, and Kant himself posits the possibility of such a revolution in the *Religion* (R 6:48).

These first two problems (ostensibly claiming the evil disposition is universal and inextirpable) stem from an apparent identification of the propensity to evil and possession of the evil supreme maxim. Seiriol Morgan's rational reconstruction solves these problems simply by separating these two, as we shall see. A final problem is that whilst Kant thinks the propensity to evil is universal, he excuses himself from providing the formal proof of it because he takes the many "woeful examples" in the world as sufficient grounds for it (R 6:32). Morgan's reconstruction is a formal proof of the propensity to evil.

4 Morgan's reconstruction: the will's reasons qua free will

Morgan's position is that the possession of the evil supreme maxim and the propensity to evil are two different things. The former is equivalent to being an evil person – to being fundamentally committed to prioritizing self-love over duty. The latter is a universal and inextirpable *incentive* to choosing the evil supreme maxim. It is the *mere temptation* to choose and thereby be committed to evil rather than that commitment itself.

The strategy is to show that the will can have reasons for choosing one supreme maxim rather than the other and that these reasons lie in its own nature as freedom. The thought is that since the will has no maxim higher than a supreme maxim to guide it in its choice of supreme maxim, only its freedom – its sheer spontaneity – can provide it with a reason. Given that true freedom is autonomy, only that maxim which enjoins autonomy can affirm the will's freedom. However, some

wills choose evil. And given that the affirmation of freedom is the criterion of choice, these wills must have been seduced by some sort of false freedom which the will offers itself. This freedom Morgan calls license, and the offering of it to oneself is what Morgan takes to be the propensity to evil.

Morgan's first step in presenting his rationally reconstructed propensity to evil is to show that the will can have reasons simply qua free will. He has two versions of a similar argument for this. The first draws on materials from Section III of *Groundwork of the Metaphysics of Morals*. At the start of *Groundwork* III, Kant says that, as a kind of causality, the will must be subject to a law on pain of being an absurdity (G 4:446). Its 'choices' would be random and hence, not choices at all. There are only two such laws: that of natural necessity and the law of freedom (the categorical imperative). Kant claims that the law of a free will is the categorical imperative. He reasons as follows: we can adopt *either* a standpoint of passive reception in sensibility or an intelligible one of active production of ideas, but it is the standpoint of the intelligible world from which we exercise our causality, one which is independent of natural necessity. If we were determined by the causal law of natural necessity, then we would not be spontaneous. The only way to be truly spontaneous is to follow one's own law. The only such law is the categorical imperative because it enjoins us to will in a way which borrows no incentive alien to the will for its willing. However, because the law of natural necessity would allow us to avoid absurdity as well, and because we can adopt the standpoint of the sensible world, one might wonder why we cannot take the law of natural necessity as our law. The answer is that as rational beings we belong more fundamentally to the intelligible world because causality is not a feature of the world itself. Rather, it is a pure concept of the understanding through which *we* synthesize the manifold of experience. Causality is dependent on the spontaneity of our minds. We can conclude that our independence from natural necessity as spontaneous beings and the priority of the intelligible world commit us to our own law – the categorical imperative. In other words, our spontaneity (our freedom) gives us a reason for choice and, moreover, an overriding reason for choosing morality.

As mentioned above, Morgan also represents a second argument given by Christine Korsgaard (1996) which reaches the same conclusion. Korsgaard uses a conceit in which we imagine the will trying to choose its supreme maxim 'before' it 'enters the world'. The following constraints apply in this thought experiment: firstly, this is a free choice, so we must exclude the idea of a supreme maxim imposed by nature. Second, if we

adhere to Kant's strictures, there is only a choice between the moral or the evil supreme maxim. Third, there is no appeal to morality or self-love available as that would issue in circular justification. Fourth, the will cannot just 'plump' for one or the other because the supreme maxim would not be reason-giving for the choice of maxims lower down in the hierarchy. A choice of self-love would limit its willing to following happiness through the satisfaction of inclinations. These are given to us by nature and their satisfaction consists in acting as though we are unfree, (i.e. as though we are part of the causal mechanism of nature). Such a choice would be an "abrogation of freedom" (Morgan 2005: 78). This will, with nothing else to guide its choice, only has its "sheer power of choice, the will's spontaneity" (Morgan 2005: 77) for this purpose, so the will has overriding reason to choose the moral supreme maxim. Once again, the two key results of this first stage of the argument are that (1) the will can have reasons qua free will and (2) the freedom of the will gives it overriding reason to choose morality.

Next comes the claim that we know that some wills choose evil[3] and the question of why they would do this despite the overridingness of morality. The notion that the will can have reasons qua free will provides a starting point in answering this. As we have seen, the promise of the affirmation of its freedom provides the will with a reason for choice. So if evil seemed to promise such an affirmation, it would seem choice-worthy. This is precisely Morgan's solution to the puzzle. He argues that the will offers itself a kind of pseudo-freedom masquerading as true freedom, the acceptance of which offer issues in that will's being evil. This false freedom Morgan calls *license* and is the unrestricted pursuit of what Kant calls outer freedom. Outer freedom is freedom from external restrictions on action in the phenomenal world. Because I cannot exercise my spontaneity without outer freedom, it may seem that the more outer freedom I have, the freer I am *simpliciter*. But actually, exercising outer freedom to the extent that one tramples on the rights of others is wrong and hence an abrogation of true freedom, i.e. autonomy.

Morgan's idea of the will tempting itself with the offer of the false freedom of license is identified with Kant's propensity to evil but *not* with Kant's idea of the possession of an evil supreme maxim. Under Morgan's reconstruction, rather than a settled policy, the propensity is now a mere

[3] This, I presume, is because unlike moral acts whose worth can be uncertain, some evil acts are clearly evil and so serious that they could only have been carried out by a fundamentally evil person – by someone possessed of the evil supreme maxim.

susceptibility to evil to which the good may fall foul and to which the evil already have. On this account, one adopts the evil supreme maxim (arrives at that settled policy) if and only if one endorses the incentive to license. This, I would submit, is an improvement on the Kantian picture whose theoretical resources cannot distinguish between these two things, at least, not in any obvious way. In addition, as mentioned earlier, since the two are seen as distinct, Morgan is able to posit on Kant's behalf a universal propensity to evil which does not also saddle the latter with the unintuitive claim that we are all evil. Also, this separation of propensity to evil and evil supreme maxim avoids the problem of an inextirpable propensity making the evil supreme maxim inextirpable.

Finally, the reconstruction provides the transcendental deduction of radical evil missing from the *Religion* because it explains the conditions of the possibility of wrongdoing in the phenomenal world. The reason any will chooses anything must ultimately lie in the overriding normativity of freedom. The only way evil could be chosen is on the basis of the (false) promise of the affirmation of freedom. So all evil wills must have accepted licentious freedom as an incentive. Since the incentive arises from the will's nature as spontaneity, which is a feature shared by all wills, the incentive is present in all wills. This is a formal proof and hence underwrites a universal propensity to evil (unlike Kant's "many woeful examples" of wrongdoing).

5 Self-deception regarding licentious freedom

However, the will knows that its autonomy gives it overriding reason to endorse morality, so how does a will come to endorse the false freedom of license and thereby adopt the evil supreme maxim? It cannot be by accident or through ignorance: in both those cases, it would not be a *choice*, i.e. something for which one is responsible. The only way evil could be chosen is if the will deceives itself into supposing that the false freedom of untrammelled license is freedom *simpliciter*. However, self-deception has its own problems. The first of these I wish to call the Belief Paradox. It seems that at least *prima facie*, self-deception involves the paradox that the self-deceiver both knows or at least believes some proposition *p* and believes that not-*p* at the same time. In the case of the Kantian will, *p* might be something along the lines of "My nature as freedom consists in and can only be affirmed by a policy of autonomy".

The second problem I call the Deceiver Paradox. We saw that the will must knowingly *choose* evil to be responsible for it. By the same token, the act or process of self-deception which facilitates that choice must

be deliberately chosen and something for which the evil will is responsible. But this, potentially at least, generates the paradox that one knows about the deception as deceiver and does not know about it as dupe. Kant himself mentions his own perplexity regarding the way one is able successfully to lie to oneself:

> It is easy to show that the human being is actually guilty of many inner lies, but it seems more difficult to explain how they are possible; for a lie requires a second person whom one intends to deceive, whereas to deceive oneself on purpose seems to contain a contradiction. (MS 6:430)

It is my view that Jonathan Webber's interpretation of Sartrean bad faith dissolves the two paradoxes whilst preserving agential responsibility as Kant's strictures require. Before examining how this is achieved, it may be useful to explain that interpretation.

6 Interpreting bad faith

Sartre's chapter on bad faith in *Being and Nothingness* is very open to interpretation and it has prompted a good deal of discussion amongst his commentators about, amongst other things, what an agent is actually aiming to do when they engage in this project. I make use of Jonathan Webber's interpretation because it yields a theory of self-deception which dissolves the paradoxes described above, but which is also particularly well-suited to application to the case of a Kantian will as Sartre's theory sees the act of self-deception as intentional (and everything a Kantian will does must be intentional).[4] Let us turn to the account.

Sartre believes that bad faith is a project whose aim is to alleviate a certain type of anxiety about our freedom.[5] According to him, each of

[4] I should say that I am not especially interested in how defensible this account is as a piece of exegesis – it may or may not be more accurate than other renditions of Sartrean bad faith. As far as the present purpose is concerned, its value lies in its ability to dissolve the paradoxes of self-deception whilst seeing self-deception as an intentional project.

[5] It should be noted that the *aim* of Sartrean bad faith – the mitigation of anxiety over the use of freedom – is of no concern to us as regards the application of this theory of self-deception to the case of the Kantian will choosing evil. I delve into this issue of anxiety now merely as part of an explanation of Sartre's conception of self-deception (or rather, Webber's interpretation of it). In borrowing from Sartrean bad faith, I only wish to make use of the *structure* of this particular form of self-deception, not its content or its particular aims.

us chooses a certain set of projects. Our character traits consist in these projects and give rise to our seeing the world as making certain demands upon us. The anxiety we feel concerns these apparent demands and our responses to them. It is because the apparent demands are based on projects which are freely chosen that the anxiety we feel about them and our responses to them amount to an anxiety about our freedom. Since we are all free, this anxiety is a feeling universally felt and bad faith, a project universally undertaken.

Before we look at what bad faith is, it is necessary to explain some terms whose use is peculiar to Sartre. In his philosophy, there are two aspects of human existence: *facticity* and *transcendence*. Facticity seems to include one's body, environment and the history of one's freedom. Webber (2009: 76) argues that it also should be taken to include one's present character, which is determined by the projects one has freely chosen. Transcendence seems to be the power to change one's character – i.e. one's projects.

Webber posits three varieties of what he calls "bad faith in the broad sense".[6] The first he identifies with the Sartrean notion of *sincerity* and involves taking properties one actually possesses as fixed (whereas the truth is we are free to change them). The other two are both forms of what Webber calls "bad faith in the narrow sense". In both of these, one denies unwanted properties one has. In the first form, this is done by pretending to have properties other than those one actually possesses and by also taking them to be fixed. In the second form, one denies unwanted properties by focusing on other properties which one does in fact possess and, again, these are taken to be fixed. Webber's contention is that by regarding these properties as fixed, one mitigates the sense of anxiety brought about by the knowledge that they are freely chosen. It is the first form of bad faith in the narrow sense which I will later apply to the case of the Kantian will deceiving itself into a choice of evil and which therefore concerns us most, presently. One instance of it can be found in Sartre's own example of a woman on a date.

In the example,[7] the woman's companion is ostensibly polite and charming, although his real project is to sleep with her. However, she, for the moment at least, wishes to take his compliments at face value

[6] As two of these do not suit my purposes in relation to the Kantian will, I will forego explaining them here.

[7] The example, arguably, has two parts, and in the second, the woman exhibits the second form of Webber's bad faith in the narrow sense when she denies the property of her sexuality by emphasizing (what she takes to be fixed) properties of sentimentality and intellect.

and enjoy his company. Her complex requirement of his desire is that first it address "her full freedom" since "the desire cruel and naked would humiliate and horrify her" (Sartre 1957[1943]: 55). And yet she also wishes that his desire "address itself to her body as object" because "she would find no charm in a respect which would only be respect" (Sartre 1957[1943]: 55). It seems that the woman wants to be admired sexually but does not want to admit this to herself in full consciousness. Correlatively, she does not want to admit to herself that the attention she is enjoying is sexually charged.

I think Sartrean bad faith dissolves the Belief Paradox by presenting us with something which plausibly counts as self-deception, but which does not posit an agent who both believes the true but unpleasant fact and at the same time believes its agreeable negation – at least not one who straightforwardly does this. Instead, Sartre's agent knows the true and unpleasant notion. Sartre says, "I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet I must know the truth very exactly *in order* to conceal it" (Sartre 1957[1943]: 49). In the case of Sartre's agent, this notion is that her character and the demands the world appears to make of her are the result of her freely chosen projects.

However, the Belief Paradox is avoided because, I would argue, it is not the case that the agent also (straightforwardly) *believes* the opposite:[8] i.e. it is not the case that she *believes* that it is not the case that the demands the world appears to make of her[9] are the result of her freely chosen projects. If she actually fully believed this, arguably she would have no cause for anxiety which she does in fact have. Instead of actually *believing* the false but pleasant notion, the agent pretends that it is true by distracting herself from the true but unpleasant thought by wilfully misinterpreting evidence – by pretending that it shows what she wants it to show. In the case of the 'coquette', we can see this idea of pretence and the deliberate misinterpretation of her companion's words

[8] At least, there is nothing, as far as I can tell, constraining Sartre or his commentators to claim that the agent must *believe* the false but pleasant notion (i.e. that they have a fixed nature) even though they use the words 'belief' and 'believe' in respect of an agent's epistemic relation to the false notion.

[9] Judging by this example and another of Sartre's in *Being and Nothingness* – a man who is under pressure from a friend to admit his homosexuality – Sartre seems to regard sexual desire as a choice. Obviously there are good grounds for regarding this as false, but we may overlook this for the sake of understanding bad faith.

in the first part of the example, in which he is ostensibly being merely innocently charming:

> If he says to her "I find you so attractive" she disarms this phrase of its sexual background; she attaches to the conversation and to the behavior of the speaker, the immediate meanings which she *imagines* as objective qualities. (Sartre 1957[1943]: 55; my emphasis)

Commentators may refer to an agent's bad faith *beliefs*, but they ought not to be taken as such. And in the following passage, Sartre (attempts to) explain the attenuated sense in which bad faith is belief (if it is to be regarded as belief at all):

> The true problem of bad faith stems evidently from the fact that bad faith is *faith*. It can not be either a cynical lie or certainty – if certainty is the intuitive possession of the object. But if we take belief as meaning the adherence of being to its object *when the object is not given or is given indistinctly*, then bad faith is belief. (Sartre 1957[1943]: 67; my emphasis)

This is hardly a conception of belief in which an agent sincerely bases what they take to be true on evidence they sincerely take to be clear and unambiguous. It also seems to require a capacity for *fiction,* which plays no part in belief in a more orthodox sense. So, it is not the case that the agent both believes *p* and believes not-*p*, and so the account does not fall foul of the Belief Paradox.

How does bad faith avoid the Deceiver Paradox? The project of bad faith involves taking evidence for one thing as evidence for another. For example, the woman takes the man's words as evidence of a polite nature. She does this by exploiting the under-determination of belief by the evidence. Since the (correct) conclusion (that the man wants sex) is not strictly required by the evidence, it leaves open the opportunity to suppose that her own conclusion is warranted. This is a poor approach to reasoning, so the *method* (as well as the evidence) is unpersuasive. This unpersuasiveness is a mark of a self-deception process and, on the face of it, something of a giveaway, and yet Sartre insists that bad faith is intentionally brought about.

Webber suggests three factors which mitigate against this ruining the deception. First, if the agent already accepts unpersuasive evidence, then *within the confines* of an attitude which accepts non-persuasiveness, such an approach will not be problematic. Second, if an action is constituted

by subordinate actions (as self-deception is), one may carry out those subordinate actions without explicitly thinking about the overall intention; to borrow an example from Webber: I do not have to think "I am walking to work" when I cross roads or avoid other pedestrians and so on. Similarly, it is not necessary to think "I am putting myself in bad faith" when I wilfully misinterpret evidence. Third, bad faith is a project and projects alter the way we see the world and ourselves: Sartre likens being in bad faith to being in a dream. Real events cannot penetrate the dream *as real events* because "the real world is no part of the dream" (Webber 2009: 101). Even those thoughts which constitute the pursuit of bad faith will seem to emanate from one's fixed nature rather than from the project as they in fact do.

7 Application of bad faith to Kant's doctrine of evil

Let us now turn to the account of how a Kantian will may use an approach to self-deception structurally similar to that of Sartrean bad faith to deceive itself into a choice of the false freedom of license rather than the true freedom of autonomy. There are a number of things a Kantian will must be able to do in order to deceive itself with regard to freedom, and there may be doubts about whether it is capable of these. Some of these worries arise because accounts of self-deception unsurprisingly deal with that affliction in, as it were, 'whole' human beings, yet I am proposing to apply such an account to the rather more rarefied entity of the human will. A self-deception story cannot be deemed suitable if it demands that its subject be able to do things (e.g. have a capacity for belief) in order to achieve a state of self-deception which that subject cannot do. We will see that the Kantian will has each capacity (or at least an analogue of it) required for self-deception.

First, we have seen that, according to Sartre, a self-deceiver must know the unwanted truth. We may wonder whether a Kantian will 'abstracted', as it were, from its human host is capable of knowledge. That the will is practical reason is perhaps reason enough to suppose that it can, in some sense, know. The will is not to be thought of as some sort of subpersonal agent or homunculus but rather simply as the person considered merely in her free and rational aspect. But, in addition to this consideration, we ought to bear in mind that some of the most important features of the practical philosophy depend on the will's being able to know. For example, it could not be a faculty of choice if it could not know what it was choosing. Nor could it make use of its higher order maxims to guide it in its choice of lower-order ones if it were not in some way aware of what

they recommended. Finally, some form of knowledge capability must be ascribed to the will if the agent is to be held responsible for the choices of maxims he makes through it. It seems that the will must be capable of knowing the unwanted fact (that its true freedom lies in autonomy).

However, Sartre's model of self-deception also requires that this knowledge be concealed or made inexplicit. Whether the will is capable of doing this is not an issue Kant explicitly addresses in any of the published works. However, there seems to be nothing in Kant's philosophy that would preclude it and, moreover, once again, there are certain features of the practical philosophy that suggest it is possible, perhaps even necessary, in some cases. For example, we have seen that the will must in some sense know what maxims it has chosen. However, the following considerations might suggest that this consciousness of what is known might be 'dim', as it were, (i.e. not foregrounded but potentially available to full consciousness through reflection). First, it seems unlikely that all of the maxims the will chooses will be in harmony with one another. It does not seem outlandish to suggest that we can sometimes glean a person's intentions (perhaps over a long period of time) and the fact that those intentions are sometimes incompatible. So, a typically fastidious will, will have at least a few inconsistent maxims. But if it were the case that the will's knowledge of its maxims always had to be fully conscious (explicit and foregrounded), then (being practical *reason*) it would not tolerate any inconsistencies: arguably, to consciously undermine one intention with another means that one or both of them are not genuinely willed. Finally, the fact (if it is a fact) that some inconsistency occurs shows that the will is not fully conscious of all of its maxims – they are not all foregrounded – even though it must know them all. In short, if the will knows all of its maxims and if it always addresses any inconsistencies amongst them of which it is fully conscious, then its knowledge of (at least some of) any inconsistent maxims it happens to have must be inexplicit. The only other possibilities are to claim that there are no inconsistencies amongst its total stock of maxims, or that it consciously tolerates inconsistencies, or that it does not know all its maxims. These all seem less plausible than ascribing dim or inexplicit knowledge or some analogue of it to the will. So, the fact that our candidate self-deception account takes this sort of inexplicit knowledge of the unwanted proposition to part-constitute self-deception is no obstacle to the incorporation of that account into the practical philosophy. When we apply this part of the account to Kant, we say that the self-deceived will *merely inexplicitly* or *dimly* knows that its true nature as freedom can only be expressed through autonomy.

Next, according to Sartre's theory, the self-deceiver must be able to believe, in some attenuated sense, the false but congenial notion – it must be capable of the faith of bad faith. Borrowing Kent Bach's[10] terminology, what seems to be required is the capacity to think *of* not-*p*, in order to avoid thinking *that p*, where *p* is a true but unwanted proposition. In the case of the will, this means thinking of the notion that its freedom consists in license to avoid thinking that its freedom consists in autonomy. Unsurprisingly, there is nothing in the corpus to suggest either that the will can or cannot block one thought from full consciousness with another in this way. However, if the will as practical reason is capable of knowledge or an analogue thereof (as I have argued above) and is therefore capable of belief, there seems to be nothing about the will which would prevent its being capable of this sort of distracting thought (the thought of not-*p*). Applying this part of the account of self-deception to the practical philosophy, we say that the self-deceived will consciously avows or *thinks of* the notion that its nature as freedom consists in the unrestrained pursuit of outer freedom.

According to Sartre, self-deception is based upon (a biased view of) the available *evidence*. If we are to deploy this account, there must be something which corresponds to evidence in the case of the will attempting to make a choice of a supreme maxim based on a conception of itself as freedom. I take it that the two conceptions of freedom: autonomy (the true conception, which demands morality) and license (the false conception, which recommends evil) must constitute the relevant evidence. It might be objected that the notion of evidence is something belonging to the empirical world, and yet I am claiming that the will can make use of it. In response, I would say that we must again acknowledge that we are attempting to deploy a theory of *human* self-deception to the abstract notion of the Kantian rational will (the person considered only in his free and rational aspect), and that if it is true that the will must be self-deceived to be evil and that human self-deception is based on (the abuse of) evidence and that the will is not the sort of thing that responds to evidence in the standard sense, then the self-deceiving will must be responding to some *analogue* of evidence.[11] I submit that this analogue

[10] These were notions he used in his papers, 'An Analysis of Self-Deception' (1981) and 'More on Self-Deception: Reply to Hellman' (1985).

[11] Perhaps talk in terms of *grounds* is less objectionable than *evidence*: we might say that the will finds grounds for taking its freedom to consist in autonomy in the true representation of its freedom as autonomy, and that it finds grounds for taking its freedom to consist in license in the false representation of its freedom as license.

consists simply in the will's presentation to itself of the two relevant conceptions of freedom (mentioned above) upon which it relies to form its self-conception as freedom.

Sartre's account is well-suited to illuminate how the Kantian will may exploit the evidence concerning its conception of itself as freedom. As we saw in the exposition of Sartre's account, the agent's bad faith is based on the exploitation of the under-determination of belief by evidence. This strategy works in different ways depending on the type of bad faith involved. From amongst the various types of bad faith alluded to earlier, perhaps the first of the two types of "bad faith in the narrow sense" in which the agent denies properties which they possess by pretending to have different ones provides the best analogy of self-deception in the Kantian will. In this type, the strategy of exploiting under-determination of belief by evidence involves making use of similarities between what the evidence actually shows and what the agent wants it to show, just as the young coquette did in taking the man's pleasant words as acts of kindness and respect rather than as part of a seduction ploy. In the same way that this woman can exploit the similarity between kind words and seductive ones, the will can exploit the similarity between freedom correctly represented as acting on one's own law (as one does in autonomy) and freedom misrepresented as acting on one's own behalf (as one does in license) in order to misconstrue it in this latter way.

The next difficulty to consider is as follows: to deceive itself with regard to freedom, the will requires a maxim (as it does for any action). Since this maxim of self-deception facilitates the adoption of the evil supreme maxim, it must also be evil. Now, one might object that ordinary evil maxims can be adopted by a good will, but the difference here is that this maxim of self-deception is *foundational* to an *overarching* policy of evil. So, it seems a will must be evil prior to adopting this maxim of self-deception. But it has also already been argued that evil requires self-deception regarding freedom. This then gives rise to a 'catch-22' type problem: one cannot be evil unless one is already self-deceived with regard to freedom, and one cannot be self-deceived with regard to freedom unless one is already evil. The solution is to claim that the affirmation of the evil desire to deceive oneself, the state of self-deception, and the adoption of the evil supreme maxim must all be 'equiprimordial'. In choosing evil, the will also chooses the self-deception required to be evil at one (literally, fell) swoop.

However, the question remains why the will should accept this evil 'package'. The answer is that this choice is ultimately unintelligible. This

may seem unsatisfactory. However, Kant says *just this* of the choice of evil (when discussing the possibility of the revolution in the *Religion*). He says, "the fall from good into evil (if we seriously consider that evil originates from freedom) is no more comprehensible than the ascent from evil back into good" (R 6:45). Also, one point Morgan makes in this regard is that since morality is overriding, any putative competing reason (e.g. evil) is no reason at all. This means we, as theorists, have a principled reason to believe that any such choice *must* be ultimately unintelligible.

Finally, we can see how the Kantian will overcomes the Deceiver Paradox. Earlier, we saw that, for Sartre, the process of bad faith involves exploiting the under-determination of belief by the evidence and that this is an unconvincing method. Sartre's agent avoids the Deceiver Paradox because he is engaged in a project of bad faith in the context of which reality cannot penetrate, not even his engagement in the process of self-deception. In the context of evil, it will seem acceptable to the Kantian will to protect that which allows it to be a 'free causality' *as it sees it*, namely its current self-conception as that which is free to will without constraint. The evil will is dimly aware of its property of autonomy but it will interpret this not as freedom but as a threat to and an unwarranted *constraint* on its affirming its current self-conception (as 'freedom'). A maxim which fends off the threat from autonomy (which seems like *unfreedom* to the evil will) is therefore acceptable in the context of evil.

For Kant, the supreme maxim constitutes what he calls the agent's *Denkungsart* or 'way of thinking'. This is the person's moral attitude towards the choice of particular maxims of action. I propose a thickening of the notion of the evil *Denkungsart*, so that in addition to being a misguided attitude about what to value, it also contains, as an ineliminable part, a wilful refusal to undo this attitude. This underscores the way in which evil in Kant should be seen as a *commitment*. The bad individual, whilst not completely incorrigible, is nevertheless very resistant to change because they have made morality something that does not immediately seem choice-worthy. Self-deception is then the first obstacle to be overcome by the evil Kantian agent who wills the adoption of morality. These considerations also highlight the way that good is the opposite of evil not just in the obvious way – as a moral direction – but also in the way that it, as autonomy, seeks to keep itself *free* from a self-induced reverie which would enslave it to the constraints of empirical desire.

References

Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Bach, K. (1981) "An Analysis of Self-Deception", *Philosophy and Phenomenonological Research,* 41(3): 351–70.

—— (1985) "More on Self-Deception: Reply to Hellman", *Philosophy and Phenomenonological Research,* 45(4): 611–14.

Kant, I. (1996) "Critique of Practical Reason", in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Ed. and Tr. M. Gregor. New York: Cambridge University Press, 133–271.

—— (1996) "Groundwork of the Metaphysics of Morals", in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Ed. and Tr. M. Gregor. New York: Cambridge University Press, 37–108.

—— (1996) "Religion Within the Boundaries of Mere Reason", in *The Cambridge Edition of the Works of Immanuel Kant: Religion and Rational Theology*. Ed. A. W. Wood. Tr. G. di Giovanni. New York: Cambridge University Press: 39–215.

—— (1996) "Metaphysics of Morals", in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Ed. and Tr. M. Gregor. New York: Cambridge University Press: 353–603.

Korsgaard, C. M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.

Morgan, S. (2005) "The Missing Formal Proof of Humanity's Radical Evil in Kant's *Religion*", *Philosophical Review,* 114(1): 63–114.

Sartre, J. (1957[1943]) *Being and Nothingness: An Essay on Phenomenonological Ontology*. Tr. H.E. Barnes. London: Methuen & Co. Ltd.

Webber, J. (2009) *The Existentialism of Jean-Paul Sartre*. London: Routledge.

9
The Pursuit of Happiness

Michelle R. Darnell

The relation between Sartre's own treatment of ethics and Kant's, or perhaps a Kantian, ethics is discussed in literature,[1] and the goal of this chapter is not to present an argument for any specific position with respect to whether Sartre's ethics are, or are not, Kantian. Nor is there a suggestion that Kantian and Sartrean ethics can be compressed into a single ethics. Rather, this chapter begins with recognition of the clear tension that exists between these two approaches to ethics, and instead of trying to resolve that tension, encourages an exploration of this tension from a new perspective. Specifically, the tension between Kantian and Sartrean ethics is developed to include direct consideration of the pursuit of happiness, which in turn is used to suggest a more complete understanding of what it means to live a 'good' life. The chapter begins with a review of major criticisms that Sartre raises against Kantian ethics, and then offers a response to these criticisms through the consideration of the role of happiness in ethics for Kant. As a result of these responses, a new charge, that Sartre does not embrace the meaningfulness of a positive account of happiness, might be raised. It is argued that Sartre's account of authenticity, and the accompanying joy that is experienced in authenticity, might be developed by more directly including happiness as a creative act that recovers the fullness or 'totality' of our being. Ultimately, it is suggested that, even for Sartre, perhaps what we might 'hope now' is happiness.

[1] Sartre himself notes the influence that Kant had on his own approach to ethics (Sartre 1956[1943]: 431). Most secondary literature on Sartre's ethics include some consideration of this influence. More extensive and explicit treatments of the relationships between Sartre's and Kant's ethics include Baiasu (2011) and Linsenbard (2007).

1 Sartre's critique of Kant's ethics

The relationship between a Sartrean and a Kantian ethics, and the influence Kant's writing had on Sartre, have been described in multiple ways, including "flirtatious" (Painter 1999), and an influence of "anxiety" (Baiasu 2003). Most simply, one may state that Sartre's criticisms of Kant are complex. Sartre, for example, praises Kant early in his career, writing that "the Kantian morality is the first great ethical system which substitutes doing for being as the supreme value of action" (Sartre 1956[1943]: 558). Immediately following this praise, however, Sartre cites a concrete, historical situation in which different parties maintain very distinct actions as morally obligatory, and suggests that ontology should be able to provide insight into what *particular* doing is "right" (Sartre 1956[1943]: 558). While this subsequent reference to particularity is not directed explicitly at Kant, Sartre's later writings suggest that this reference is likely at least in part indicative of his rejection of an entirely Kantian ethics. In *Existentialism and Human Emotions*, Sartre claims that Kant "believes that the formal and the universal are enough to constitute an ethics. We, on the other hand, think that principles which are too abstract run aground in trying to decide action" (Sartre 1998: 47). Ultimately, we may summarize Sartre's rejection of Kant's ethics on three related grounds: Kant (1) ignores concrete facticity, (2) grounds ethics in the noumenal realm, and therefore (3) does not provide a positive ethics.

While credited as an early attempt to emphasize 'doing', Sartre charges Kant's ethics as incapable of fully being committed to 'doing' because it is disengaged from the concrete world. This disengagement itself has two dimensions: according to Baiasu, it cannot account for concrete circumstance, and it cannot account for the concrete characteristics of a person (2003: 25–6). These two aspects of disengagement are certainly related, and are sometimes treated as a single problem, however it is helpful to consider them as distinct. With respect to the former, the problem is a reliance on abstract moral principles. Reliance on abstract moral principles, it is charged, results in an ethics that is "purely formal and contentless" (Anderson 1993: 51) and thereby "is deprived of any practical force" (Baiasu 2003: 25). Since to act is to "arrange means in view of an end" and "modify the shape of the world" (Sartre 1956[1943]: 559), yet abstract moral principles are contentless, i.e. not pertaining to the world, it is entirely unclear how any actions are actually ordered by the moral principles provided in Kant's ethics. Furthermore, the end we have in mind when pursuing abstract morality must not be in this

world, since it is empty of worldly meaning, and the end goal of ethics is possible only in a transcendent realm (Anderson 1993: 51). In this way, abstract morality is not meaningfully a human affair, given Sartre's ontology (Linsenbard 2007: 66).

Sartre's ontology not only rejects a separation of ethics from the lived world, it also emphasizes that human existence is always the existence of particular individuals. As such, with respect to the second dimension of disengagement mentioned in the above paragraph, the problem with Kant's ethics is an assumption of the sameness of individual persons. Baiasu describes Sartre's critique of Kantian morality as one in which "all individuals were the same abstract person" and therefore "ethical imperatives are transformed into precepts addressed to the same abstract persons" (Baiasu 2011: 121). Abstract moral principles only tell us that 'one ought to do (x)' or more precisely, 'one ought not do (y)', but then we must ask "to whom is the ethical demand addressed? To the abstract universal? But then it loses all its meaning" (Sartre 1992: 7). Sartre emphasizes that ethics is about doing, but what ought to be done is itself informed by being. The question is not 'what ought one do', but, more accurately, 'what ought I do in this situation', the answer to which is informed by the individual's concrete existence in the world as for-itself.

The criticism that the concrete existence of the individual is ignored generally concerns the denial that ethics always is grounded in situation. Posited independently of a situation, the ends of actions are *entirely given* via moral principles directed at the abstract 'one' as absolute and unconditioned (Sartre 1992: 254). But it is with the 'given-ness' of moral imperatives that a new problem is presented. Moral actions become infused with a value that is not freely chosen by the individual agent. For Sartre, this amounts to inverting value, such that value exists and thereby a demand is made against me (in violence), rather than value arising as a result of a demand I choose to take up for myself.

The charge that values are given to me in Kant's ethics gives rise to a second element of Sartre's criticism, namely that, for Kant, values can be given to me because they are grounded in the noumenal realm.[2] Furthermore, Sartre seems to assume that the noumenal is otherworldly;[3] the noumenal freedom that must ground these given values is

[2] Sartre writes that "This freedom that for Kant upholds the categorical imperative is noumenal, therefore the freedom of *another*" (Sartre 1992: 139).

[3] The suggestion here is that Sartre takes a "two worlds" approach to the noumenal/phenomena distinction, as opposed to a "two perspective" approach to the same distinction.

"freedom in me, not as I am freedom or exist it, but as it is" (Sartre 1992: 254). This, of course, leads to an ethics of bad faith within Sartre's philosophy: "I am in bad faith if I declare that I am bound to uphold certain values, because it is a contradiction to embrace these values while at the same time affirming that I am bound by them" (Sartre 2007: 48). To be clear, the problem suggested here is not that values have an element of otherness or objectivity; the problem is that abstract morality rejects the role of the individual in establishing value. For Sartre, "(value) can be revealed only to an *active* freedom which makes it exist as value by the sole fact *of recognizing it as such*" (Sartre 1956[1943]: 76, italics added),[4] and it is the individual that serves as the foundation of ethics. Sartre's concern is that, by grounding ethics in the noumenal realm, the role of the individual is diminished to merely maintaining values, rather than participating in the constitution and creation of values.[5]

If value is given to me, and therefore I need not bring it into existence, a negative ethics is offered. To be ethical, I need not actively establish value, I merely need to restrain my will from making choices or engaging in actions that limit or restrict a pre-established value. One only need *not* do *x* in order for ethics to be realized, rather than positively doing *y* in order to create value. For this reason, Kantian ethics cannot tell us what *to do* in a situation, for it becomes more about not acting against given values. As such, we are given, via the Formula of Universal Law, only negative (direct) duties. Thus Sartre: "'Do Not Lie,' (Kant) says, means 'in no case no matter the situation.' In other words the world is inessential. Let us also add: my life, my projects, my desires" (Sartre 1992: 238).

Sartre argues against Kant by calling for a positive ethics, which is grounded in the interdependency of the individual and the meaning-filled world.[6] Accordingly, ontological commitments like those of Sartre

---

[4] Note that Sartre claims value is *revealed* to me, and that I *recognize* it. But for this to be possible, value must in a sense be presented to me (exist independently of me), yet at the same time value cannot simply be imposed on me – I must make it my own. This is perhaps connected to Sartre's discussion of obligations being both one with my freedom (taken up and upheld by me) and yet distinguished from my freedom (Sartre 1992: 254). This is deserving of considerable more treatment, but is well beyond the scope of this chapter.

[5] "My freedom is no longer constituting and creative, but rather realizing. It no longer has its task to bring about the world of ethics, but just to maintain it" (Sartre 1992: 257).

[6] "the relation of interdependence established by this absolute consciousness between the me and the World is sufficient for the me to appear as 'endangered' before the World, for the me (indirectly and through the intermediary of states)

make a positive ethics possible. For example, above it was noted that, for Sartre, value is both revealed to me and I must actively recognize values to make them exist. To reconcile values as both revealed to, and created by, an active freedom, the being to whom value is revealed and the being who reveals must be separate, but separated by nothingness. For Sartre, this nothingness makes possible both my existence as an agent in the world and the existence of the meaningful world, and also establishes the interdependency between me and the world, which serves as a philosophical foundation for a positive ethics. In fact, denying the positive act of choosing value in context is, for Sartre, a denial of what it means to exist as human, that is, it is an expression of bad faith.[7]

2 The role of happiness in Kant's ethics: a means and an end

Three interrelated criticisms that Sartre holds against a Kantian ethics were offered above: Kantian ethics ignores concrete facticity, it is grounded in the noumenal realm, and therefore does not provide a positive ethics. Sartre's emphasis on a positive ethics does *prima facie* rule out a consistency between Kantian and Sartrean ethics, but this presumed rejection of a positive Kantian ethics stems from a narrow consideration of Kant's understanding of morality. Specifically, with most accounts of Kant's ethics there is a standard rejection of teleology, and a rejection of a meaningful role of happiness. A more concerted consideration of these two elements of Kant's ethics demonstrates the importance of context, creativity, our interdependence with other persons, and the basic need for individuals to create value in the world; a positive element of Kant's ethics is thereby suggested.

While it is clear that Sartre had a relatively strong understanding of Kant's ethics, it is not surprising that Sartre overlooked, or underestimated, the role of happiness in Kant's ethics. Indeed, many scholars of Kant are quick to dismiss happiness as, at best, unimportant for Kant's ethics, and at worst, completely undermining Kant's ethics. Wood has noted that setting happiness as a moral end would "contradict Kant's

---

to draw the whole of its content from the World. No more is needed in the way of a philosophical foundation for an ethics and a politics which are absolutely positive" (Sartre 1991: 106).

[7] "To refuse to see oneself as freely choosing *in relation to* this or that situation ... is to flee in bad faith from one's freedom" (Linsengard 2007: 68).

thesis that the setting of ends is always an act of freedom, never the result of natural necessity" (Wood 1999: 66), and Lewis White Beck suggests that Kant's argument for the plausibility of a *summum bonum* (happiness in proportion to worth) is inconsistent with the rest of his moral philosophy (1960: 245). Yet we can find very clear statements from Kant that establish a relationship between ethics and happiness: "the system of morality is ... inseparably ... bound up with that of happiness" (Kant 1965: A809/B837). Kant's inclusion of happiness in ethics is complicated not only by the ambiguity of the term 'happiness', but also by the ambiguity pertaining to the role it plays. As Wike (1994) has argued, happiness could theoretically be understood as a principle of morality (which Kant rejects), as a means to a moral end, or as a part of a moral end. Clarification on these different roles of happiness, for Kant, suggests that a Kantian ethics is significantly more 'positive' than Sartre allows.

One standard reason for dismissing the role of happiness in a Kantian ethics stems from a consideration of happiness as a principle. In this case, happiness is understood non-controversially as the sensible object of the will; accordingly, if happiness is taken as a principle of morality ('always act to promote your happiness'), the will is influenced through sensation and heteronomy results. As Baiasu has noted:

> When the will is determined by a practical principle in virtue of the fact that its purpose is desired, then it is determined by the matter of the principle. When the principle determines the will because its purpose is right, then the purpose does not primarily count as an object of desire, and Kant says that the will is determined not by the matter, but by the form of the principle. (Baiasu 2003: 28)

In this quote, Baiasu provides us a description of why happiness cannot be a principle of morality, but additionally his specific language that moral principles are to be determined "not by the matter, but by the form" highlights the Sartrean critique that Kant's ethics is very much independent of the world in which we live. Kant must reject the pursuit of happiness as a moral principle because of the focus on "matter", which is precisely the lack of connection between ethics and the world that grounds Sartre's criticisms. However, 'principle of morality' is not the only possible role of happiness in ethics, and Kant's recognition of happiness as a natural end for humans is incorporated in his ethics, not within a 'principle' or as objective end that we have a (direct) duty to pursue, but as nonetheless important. Further consideration of how

happiness accompanies ethics helps to address the general Sartrean critique that Kant is unable to provide a positive ethics.

Wike provides considerable evidence that happiness plays an important role in Kant's ethics in terms of a means to morality and as part of the highest good (*summum bonum*). Beginning with the former, Wike notes that, for Kant, "the natural world is not unrelated to the moral world" and "more attention needs to be paid to how natural ends and objects of inclination facilitate (Kantian) morality" (Wike 1994: 90). One way of focusing our attention on the relation between the natural world and moral world in Kant's philosophy is to look beyond the direct duties that are given to us through the Formula of Universal Law (which, as noted above, only give us negative duties) and re-consider the role of indirect duties in Kantian ethics. Because indirect duties, for Kant, are duties directed towards intermediate states or objects that support the pursuit of moral ends, an inclusion of indirect duties is "the positing of a point of transition between the natural and the moral realms" (Wike 1994: 105). Given its status as a natural end, happiness, for Kant, is a means to morality. That is, happiness serves as the ground of the possibility of an action whose effect is an end to morality. Being happy supports a person's attempts to be moral, and lacking happiness makes it harder to be moral (Kant 1996a: 6:388). Accordingly, there is an indirect duty to pursue sensible happiness: "to assure one's own happiness is a duty (at least indirectly)" (Kant 1996b: 4:399).

The pursuit of happiness cannot be a direct duty, as was shown above, because this leads to heteronomy. But as an indirect duty, happiness plays an important role in Kant's ethics. With the recognition of this indirect duty, there is the start of a meaningful connection between the moral world, which Sartre criticizes as grounded in the noumenal realm, and the empirical world. Additionally, there is an implicit recognition of individual persons needing to take into consideration the concrete details of a situation in order to act on this duty to pursue happiness. Consider what Wood says about the pursuit of happiness in his commentary on Kant's ethics:

> anyone's conception of happiness would clearly involve empirical details not only about the ends constituting it but about the means to those ends and the properties to be given to the use of such means. For this reason, it is hard to see how an assertoric imperative could be a priori and valid for all rational beings (since the happiness of one person may differ greatly in content from the happiness of another). (Wood 1999: 68)

The pursuit of happiness, both for oneself and supporting the pursuits of others (Kant 1996b: 4:423), is not a pursuit that is dictated to us as needing to unfold in any particular way. Even with an acceptance that all finite rational beings have happiness as a natural end, what that means for me, and the process by which this happiness might be achieved, is distinct from the meaning and process of achievement for you. Accordingly, if I am to support your pursuit of happiness, not only must I understand what that means for you, I also need to respect your freedom in establishing the meaning of (the pursuit of) happiness. In this way, there is the start of a Kantian response to Sartre's challenges that a positive ethics is not given: with regard to the pursuit of happiness, the concrete existence of the individual must be embraced, and both the means and specific ends of actions are *not* entirely given to me as absolute and unconditioned, but rather are determined in situation.

We have seen that recognizing the importance of happiness as a means to morality in a Kantian system is a partial response to Sartre's concerns, but we may conjoin this response with another, namely the need to recognize the teleological dimension of Kant's ethics. While many ignore the teleological dimension of Kant's work, a number of prominent scholars have offered strong arguments about the importance of this dimension. Consider the specific language Wike uses to describe why it is important to recognize the teleology within Kant's ethics: "principles of ethics (the form) are useless, without an application, apart from the ends or objects whose pursuit they direct" (Wike 1994: 31). This sentence could very well be (at least a partial) description of Sartre's critique of Kantian ethics: if Kant's ethics concern only the formal element of principles, they are not meaningfully related to my concrete existence. On such grounds, we could condemn Kant for not actually providing an account of *practical* reason, which concerns actions. Actions are always performed with an end in mind, and principles independent of matter are incapable of directing action. To be sure, for Kant, ethics cannot *begin* with ends (for this would be heteronomy), but ethics, qua the exercise of practical reason, does necessarily *lead* to ends. For this reason, Kant must, and does, include a teleology within his ethics: "pure practical reason is a faculty of ends generally, and for it to be indifferent to ends, that is, to take no interest in them, would therefore be a contradiction" (Kant 1996a: 6:395). This teleological dimension enables a union between form and matter, and between the exercise of pure reason and the natural, concrete world within which we live. Recognition of this teleology in Kant provides the grounds of a rebuttal to the charge that Kant's ethics are purely formal and contentless.

Wike argues that ends work in conjunction with the form of the maxim to determine the will in Kantian ethics. This conjunction must in a certain sense be unique, insofar as while the end might (in part) determine the will, heteronomy must still be avoided. Wike further accounts for this by suggesting that in an autonomous will the ends of pure reason act to (indirectly) determine the will because they follow from the will; in this manner, there is no "choosing between" a material determining ground and a formal determining ground of the will, since a choice for one is necessarily a choice for the other (Wike 1994: especially 77–80). Most generally, this conjunction is brought together in what Kant calls the final end for all humans, happiness in proportion to worth. On one interpretation of Kant's views, this happiness which forms part of the highest good is a sensible state, even if it is restricted to a sensible state that is consistent with virtue. Understood this way, happiness (in proportion to worth) is something we could hope for in the natural world. While this interpretation is not without its critics, this provides an interesting possibility of how morality and the natural world are brought together in Kantian ethics, and a possible response to the Sartrean critique that Kant's ethics is grounded in the noumenal realm alone, and is disconnected from our empirical existence.

Furthermore, Kant tells us that the final end is the unification of *all* moral ends (Kant 2001: 6:5). While again not without critics, a strong case can be made that the unification of all moral ends is not establishing the highest end merely for an individual, but is the *social* uniting of our individually willed moral ends. Guyer notes that "it is never merely one's own happiness that is aimed at in the highest good, ... the highest good itself is not a condition that may be attained by an individual – although it must be aimed at by every individual – but a condition that may be realized only by mankind as a whole" (Guyer 2000: 392).[8] Kant is somewhat explicit in claiming the end of morality is social, writing that the end is "to unite in an ethical community" (Kant 2001: 6: 151), a duty which is "not of human beings toward human beings but of the human race toward itself" (Kant 2001: 6: 97). Given this end of ethics, it is clear that Kant's ethics contains an interpersonal element. Furthermore, because this interpersonal element is the end,

[8] See also Wood's account of Friendship in Kant as "a relation in which one's happiness is effectively promoted not through one's own striving but through a common striving toward an end in which both happinesses have been included" (Wood 1999: 279–80).

the material *determining ground* of the will, in an important sense there is an interpersonal origin of ethics for Kant.

3 The role of happiness for Sartre: Bad faith

The above descriptions of the role of happiness in Kant's ethics are not without controversy, nor do they provide some sort of absolute reconciliation between Kantian and Sartrean approaches to ethics. Rather, the immediately preceding section of this chapter suggested that preliminary responses to Sartre's criticisms, that Kant's ethics do not provide a positive ethics, ignore concrete facticity, and are fundamentally 'other worldly', might be formed upon a more serious consideration of the role of happiness in Kant's ethics. Another interesting consequence of introducing the role of happiness in Kant's ethics, however, is drawing attention to the role – or missing role – of happiness in Sartre's ethics.

Just as within Kant's philosophy, happiness is not a topic that the average reader would emphasize in Sartre's philosophy. Indeed, happiness itself is mentioned in only a few passing remarks in Sartre's non-literary works, and in these remarks, it is not so much happiness, per se, that is addressed, but rather unhappiness:

> The being of human reality is suffering because it rises in being as perpetually haunted by a totality which it is without being able to be it, precisely because it could not attain the in-itself without losing itself as for-itself. Human reality therefore is by nature an *unhappy* consciousness with no possibility of surpassing its *unhappy* state. (Sartre 1956[1943]: 140, italics added)

If happiness is taken generally as the satisfaction of inclinations, the perpetual haunting of being for-itself-in-itself, which is an impossible goal, leaves human reality as an unhappy state. As indicated by the above quote, an account of (un)happy consciousness is related to an account of bad faith, for which the first act is "to flee what it is" (Sartre 1956[1943]: 115). Specifically, it follows from an account of unhappy consciousness that attempts to be happy are themselves acts of bad faith.

This thesis is supported by the treatments of happiness in Sartre's novels, plays, and even his autobiography, which consistently present the pursuit of happiness, or claims of being happy, as indicative of bad faith. Indeed, a careful study of the role of happiness in Sartre's literary works reveals various patterns of bad faith: lying to oneself or being an accomplice to others' attempts to lie to themselves; denying the

temporality of being-for-itself or the structure of temporality; embracing a deterministic attitude towards human existence; and denying being-for-itself as inclusive of being-for-others. To be sure, happiness is never a direct topic in Sartre's works of literature, but it is a theme that appears in many of his literary works. In what follows, these references to happiness are used to illustrate Sartre's association of happiness with bad faith.

One of the most cited descriptions of bad faith is a "lie to oneself" (Sartre 1956 [1943]: 89), and Sartre's inclusion of happiness, or making other people happy, is often presented as conjoined with an attempt at such a lie. In *The Words*, Sartre presents lying as a component of happiness ("a pious lie which I tell them in order to make them happy" (Sartre 1964: 33)), and in *Dirty Hands* happiness is correlated with an acceptance that the for-itself can be reduced to being-in-itself (Sartre 1976a: 131, 153, and 196). There is then the further implication, in *Dirty Hands* and also, for example, in *The Room*, that happiness is simply bodily pleasure, often brought about through sadism or masochism.[9]

Another way in which happiness is related to attempts of the for-itself to "flee what it is" (Sartre 1956 [1943]: 115) is the association of happiness with a denial of the temporality of being-for-itself. Sartre presents happiness as an attempt to re-live the past, to preserve the present as an isolated moment in time, into which one's future self can escape, to deny the present-ness of activity, and to deny the role of the past in ascribing meaning to present action. Interestingly, some of the strongest examples of happiness as a denial of the temporality of being-for-itself come from *The Words*, Sartre's so-called autobiography. In one section, Sartre recounts a regular childhood practice of re-enacting his birth for his family, a practice by which he will have "made one more person happy" (Sartre 1964: 31–2). In this scenario, happiness is found by resting in the given-ness of the past, whereby even present action is a reoccurrence of actions that have already been completed and the outcome is determined. In another reflection within *Words*, Sartre recounts when, to his dismay (he was hoping for a novel), his mother gave him a collection of questionnaires so that he might interview his friends and build "happy memories" (Sartre 1964: 107). Here, his mother is anticipating that future pursuits of happiness will be successful by being able to appeal to the past (Sartre's childhood). In both of the preceding examples, the past is

[9] See examples of the connection between happiness and physical pleasure in Sartre (1976a: 170, 177); and examples of happiness equated with physical pleasure and objectification of others in Sartre (1948: 45).

given "a kind of honorary existence" (Sartre 1956[1943]: 161), and there is an isolation of the past from the present, which enables the belief that the past can 'be again' in the present. This isolation of the past from the present is further demonstrated when Sartre recounts being a child who was "as happy as can be to see Prussia parading by to the sound of that puerile music" (Sartre 1964: 37), but obviously not understanding how the past influences the meaning or value of this present parade. That is, happiness was achieved in part because of a mistaken belief that the past is an isolated element of time, and the present is an "instantaneous island" (Sartre 1956[1943]: 161), rather than accepting that the past is organically, internally, related to the present.

More broadly, the pursuit of happiness is presented as attempts to flee from the synthesis of moments of time, and instead posit the elements of time (past, present, and future) as givens (Sartre 1956[1943]: 159). In this way, Sartre recounts being "only too happy to put (a task) off until the following day" (Sartre 1964: 116), whereby the future is set as an isolated moment in time that releases one from present responsibilities. Indeed, attempts to escape from our normal experience of temporality in sleep[10] brings "happy dreams" (Sartre 1976b: 6), but of course "one puts oneself in bad faith as one goes to sleep and one is in bad faith as one dreams. Once this mode of being has been realized, it is as difficult to get out of it as to wake oneself up; bad faith is a type of being in the world, like … dreaming" (Sartre 1956[1943]: 113).

In bad faith, the for-itself pursues happiness by fleeing from its existence as for-itself, and treating past, present and future as entirely given. If "I want to make you happy" I desire to "give you everything you want" (Sartre 1948: 60) without recognizing that the choosing and satisfaction of such wants is a temporal unfolding. This is, of course, a denial of the internal relation between the past, present and future within the temporal upsurge that is being-for-itself. Such a denial reinforces the conflation of being-for-itself with being-in-itself, such that the existence of the for-itself is solidified, with its past and future no longer in question. Happiness is then found in the complete given-ness of existence. Thus, of the characters in films Sartre would watch as a child: "How happy were those cowboys, those musketeers, these detectives; their future was there … and governed in the present" (Sartre 1964: 124).

A determined existence relinquishes one's self from the responsibility of choice and uncertainty, and it is with the escape of freedom that one apparently finds happiness. Sartre references happiness as possible

[10] "The past has slipped away from it like a dream" (Sartre 1956[1943]: 163).

because one convinces oneself of already knowing the future, or rather believing the future already *is* in the present. Sartre recounts convincing himself that "the unforeseen could only be delusion" because of having decided "in advance that my story would have a happy ending" (Sartre 1964: 233). Even when reading a novel, we often console ourselves from the apparent injustice that befalls the (anti-)heroes we read about by not being "fooled" by what seems mere chance: "The reader is not fooled; he has leafed through the last chapter to see whether the novel has a happy ending" (Sartre 1964: 243). When we are forced to recognize that our existence is not determined, we are unhappy, but even then, a man who willingly undertakes an action often proclaims "he *would be* happy if he were prevented from carrying it out" (Sartre 1976a: 224, italics added).

Bad faith is denying one's existence as for-itself, and the implications of this are not merely attempts to posit one's existence as in-itself, or determined, but also to deny being-for-others as an ontological structure of the for-itself. In *Words*, Sartre describes several childhood situations in which he was being ignored and "barely tolerated" by others (Sartre 1964: 140). Positing one's existence as in-itself releases oneself from the struggle of being-for-others. When Sartre recalls not being included in play with other children he notes that "even a silent role would have made me happy"; here, happiness is sought by being a mere object for others (Sartre 1964: 134). Simultaneously, happiness is also sought by attempting to remove the for-others structure from the existence of being-for-itself, and Sartre writes "I remember happy convalescences and a black, red-edged notebook which I would take up" as an escape from the social reality within which he found himself (Sartre 1964: 134).

4 A positive role of happiness in Sartre?

Given the strong connection between happiness and bad faith in Sartre's philosophy, a striving for a more authentic existence seemingly calls for a rejection of the pursuit of happiness, and Sartre writes that over time he "developed a hatred of happy swoons" (Sartre 1964: 112). Obviously, then, a new source of tension between Kantian and Sartrean ethics is revealed. It was suggested above (Section 2) that a consideration of the positive role of happiness in Kant's ethics enables a means by which his abstract morality is united with concrete existence of humans; but it was also suggested (Section 3) that Sartre posits happiness precisely as a means by which one attempts to deny one's own existence as for-itself. As such, happiness serves as an interesting dimension to the

tension between Kant's and Sartre's philosophies, and raises the question of whether there is also a positive role of happiness in Sartre's philosophy.

Few scholars have addressed the role of happiness in Sartre's philosophy directly. Perhaps the most obvious example of such scholarship is found in Wang's *Aquinas and Sartre: On Freedom, Personal Identity, and the Possibility of Happiness* (2009). In this text, Wang ultimately accepts that, for Sartre, happiness in this world is an impossible ideal, but suggests that it is precisely because the achievement of happiness is impossible that we are assured of our free existence. Subsequently, we are confident in the possibility that we could strive for happiness, even if the achievement of happiness will not occur. In short, Wang argues along the lines of 'loser wins' that the impossibility of happiness itself offers some sort of optimism. Such an approach to happiness in Sartre's philosophy, however, does not admit much of a positive role of happiness. Wang does hint at the option that the impossibility of happiness might be overcome with a radical conversion into authenticity, but does not address that option in-depth. The question then remains as to whether authenticity might be related to happiness, such that happiness can take a more positive role in Sartrean ethics.

For Sartre, authenticity "transcends the dialectic of sincerity and bad faith" (Sartre 1992: 474). Because happiness is mostly presented by Sartre as an expression of bad faith, it would seem that an escape of bad faith would further entail a rejection of the pursuit of happiness. Yet, if we consider that the radical escape of bad faith occurs with a "self-recovery of being which was not previously corrupted" (Sartre 1956[1943]: 116 n 9), there might also be a radical conversion in our understanding of happiness. In bad faith, happiness is merely an attempt to not be unhappy, and the unhappy consciousness is assumed to be the only mode of existence of the for-itself. Rather than happiness as the escape from unhappiness, in authenticity perhaps it is possible to recover a meaning of happiness that is not based on what is impossible, but instead based on the affirmation of one's existence as being-for-itself. "The authentic person gives her life meaning and value by accepting and affirming herself as the free creator of a meaningful world" (Anderson 1993: 58). With this acceptance and affirmation, we are told, the anguish of freedom is converted into joy.

> The authentic individual ... wills her freedom to be the foundation of the world. [She] ... ceases to be a totally unjustified contingent existent living in a pointless universe. With joy she sees her existence

> as having a task or purpose, for it is the foundation of, and thus essential to, that meaning-filled world. (Anderson 1993: 58)

In bad faith, the pursuit of happiness was the impossible pursuit of being-for-itself-in-itself, or a totality that is paradoxically both chosen and given. In authenticity, however, my existence is justified and has purpose; in a sense, then, there is a 'totality' to my existence. This totality is conceived not as a static or necessary being (as in bad faith), but as a recovery of the fullness of my being whereby all of my free acts, taken together, are understood as the source of a meaningful world.[11] I grasp that I am not the creator of the being of the world, i.e. I am not God or the absolute for-itself-in-itself, but I am the source of the *meaningful* world. I creatively offer my being as a gift for others, and experience "aesthetic joy, an emotion which also includes feelings of security and sovereign calm, because I see my existence, not as an unjustified contingency, but as the necessary foundation and essential cause of a meaningful ... universe" (Anderson 1993: 57). In authenticity I accept, will, and take responsibility for the fullness of my existence, and with this comes joy. The joy that accompanies my embracing of the fullness of my existence might be described as 'happiness', thereby revealing a positive role of happiness in Sartrean philosophy.

Any happiness within authenticity must be a rejection of the impossible goal of being for-itself-in-itself, and an embracing of the creative *activity* of my being which is the purpose of my existence and thereby is a *revealing* of my fullness: an active "re-affirmation of ourselves and our world as a value to be *pursued*, rather than a self or world that *is*" (Linsenbard 2007: 76, italics added). In contrast, the bad faith pursuit of happiness is the pursuit of an impossible *static* totality, with the possibility to assert that I *was* happy "only in the past" (Sartre 1956[1943]: 172) and the final evaluation of one's life as happy or unhappy possible only at one's death (Sartre 1956[1943]: 169).

In addition to the conversion of happiness from a state (bad faith) to an activity (authenticity), happiness also must no longer be understood as related to *isolated* consciousness – in bad faith I try to isolate moments of time, myself from others, or my existence as entirely given

[11] This interpretation of a fullness of being in authenticity is suggested in parallel to the position in Kantian scholarship that happiness is not necessarily the satisfaction of all inclinations, but as the satisfaction of a system of inclinations. See Wike (1994: 5–9) for an account of this position in Kantian scholarship.

and independent of contingent factors of the world – but as a *social* activity. In authenticity, I offer myself and creation as a *gift* to others, and appeal to the freedom of others to accept and use my gifts as each so chooses. To gift is to will the freedom of others, "to rejoice in the Other's being-in-the-world" (Sartre 1956[1943]): 508), and to adopt each other's projects as part of one's own (Anderson 1993: 67). Here there is a teleology in Sartre's philosophy as well. Sartre posits the goal of human reality is creating a city of ends.[12] He also says, however, that the goal is not a city of ends, or *happiness*, that can be achieved *once and for all* and bring history to a close, it is a city of ends as a *becoming* – not a static end state (Sartre 1992: 169).

A teleology or setting of a final goal in Sartre's philosophy must also be an embracing of active creativity, and this is suggestive of Sartre's treatment of art in *What is Literature*.[13] In this text, Sartre writes that the final goal of artistic creation is to "recover this world by giving it to be seen as it is, but as if it had its source in human freedom" (Sartre 1993: 43), whereby the creation is both value and task. There is an interesting parallel between artistic creation and authenticity, and between the creative act and a "recovery of *the totality* of being" that is joyfully experienced (Sartre 1993: 42, italics added). The joy of authenticity, the satisfaction of recovering the totality or fullness of my being, might be understood as an engaging in the art of happiness. It is worth considering that there is a positive role of happiness in Sartre's philosophy, and that is to be artists of happiness.

5 Conclusion: on what can we hope

The possibility of a positive role of happiness in Sartre's and Kant's philosophies was presented above, but for both of these philosophers, happiness must necessarily remain a possibility. For Kant, achievement of the *summum bonum* cannot be guaranteed by pure reason, precisely because it includes the sensible element of happiness. For Sartre, if we were to accept happiness as an accompaniment to authenticity, happiness cannot be a given since it is an artistic *pursuit*. In recognition of the status of happiness as only a possibility, even a necessary possibility, Kant posits that happiness is something for which we can hope (Kant 1965: A805/B833).

---

[12] See Sartre (1956[1943]: 10, 88–9, 166, 168–9, 290, 402–3).

[13] This idea comes from Anderson (1993, esp. 57).

McBride briefly entertained the topic of "Sartre's Response to Kant's Question, 'What May I Hope'", but quickly dismisses Kant's own response to this question, namely the *summum bonum*, as a viable Sartrean response, saying it is "of course not within Sartre's reach" (McBride 1999: 59). This dismissal, however, is grounded in McBride's acceptance of the *summum bonum* as realized in eternal life; while some Kant scholars do argue the *summum bonum* is to be found in the intelligible world, this is certainly not the only interpretation of the *summum bonum*, and is not the interpretation required of the presentation above. Given the preceding sections of this chapter, Sartre's response to Kant's question might not be entirely incompatible with Kant's answer: happiness (understood, of course, in a very specific way).

Hope, like happiness, is not a major theme in Sartre's writings. His earlier writings, in fact, focus more on despair than hope, just as they did unhappiness more than happiness. In *Existentialism as a Humanism*, he writes that "I must commit myself, and then act according to the old adage: 'No hope is necessary to undertake anything'" and that this "helps people to understand that reality alone counts, and that dreams, expectations, and hopes only serve to define a man as a broken dream, aborted hopes, and futile expectations; in other words, they define him negatively, not positively" (Sartre 2007: 36–8). It is worth noting that here hope is taken as a (unnecessary) condition of action, to be distinguished from action itself. That is, hope is here presented as passive. In his essay referenced above, McBride reminds us that hope is traditionally symbolized by an anchor, a source of subsurface stability and security, and that "Sartre, ever restless ... in his view of human existence, would surely not find much value in the notion of hope thus symbolized" (McBride 1999: 65).

Yet, scholars are often quick to point out that, despite the presence of despair, existentialism is optimistic. In a sense, despair seems unavoidable given that there are contingent factors independent of my will that impact the success of my actions, but Sartre claims that those who assert existentialists are without hope are confusing two meanings of despair (Sartre 2007: 53–4). Although not explicit, the suggestion is that one sense of despair is a lack of passive hope. Hope, however, might also be understood as action. Thus, Sartre: "the only hope resides in his actions" (Sartre 2007: 40). Hope maintained in authenticity is not passive, it is active progress. Indeed, in *Hope Now*, Sartre describes one of his final "naïve" ideas to be a belief in progress, reporting "I still feel that hope is my conception of the future" (Sartre and Levy 1996: 110). For both Sartre and Kant, the future towards which we ought to

be progressing is a city of ends. Given the considerations throughout this chapter, it seems both philosophers might hope for happiness in our progress.

References

Anderson, T.C. (1993) *Sartre's Two Ethics: From Authenticity to Integral Humanity*. Chicago: Open Court Press.

Baiasu, S. (2003) "The Anxiety of Influence: Sartre's Search for an Ethics and Kant's Moral Theory", *Sartre Studies International,* 9(1): 21–53.

—— (2011) *Kant and Sartre: Re-discovering Critical Ethics*. New York: Palgrave Macmillan.

Beck, L. W. (1960) *A Commentary on Kant's Critique of Pure Reason*. Chicago: University of Chicago Press.

De Beauvoir, Simone (2000[1948]) *The Ethics of Ambiguity*. Tr. B. Frechtman. New York: Citadel Press.

Guyer, P. (2000) *Kant on Freedom, Law and Happiness*. Cambridge: Cambridge University Press.

Kant, I. (1965) *Critique of Pure Reason*. Tr. N. Kemp Smith. New York: St. Martin's Press.

—— (1996a) "The Metaphysics of Morals", in *Practical Philosophy.* Ed. and Tr. M. J. Gregor. Cambridge: Cambridge University Press.

—— (1996b) "Groundwork of the Metaphysics of Morals", in *Practical Philosophy.* Ed. and Tr. M. J. Gregor. Cambridge: Cambridge University Press.

—— (2001) "Religion Within the Boundaries of Mere Reason", in *Religion and Rational Theology*. Ed. and Tr. A. W. Wood. Cambridge: Cambridge University Press.

Linsenbard, G. (2007) "Sartre's Criticisms of Kant's Moral Philosophy", *Sartre Studies International,* 13(2): 65–85.

McBride, W. (1999) "Sartre's Response to Kant's Question, 'What May I Hope' ", *Phenomenological Inquiry,* 23: 58–73.

Painter, M. (1999) "The Profane Become Sacred: The Protestant Ethics of Kant and Sartre", *Southwest Philosophy Review,* 15(1): 211–17.

Sartre, J.-P. (1948) *The Room* in *The Wall, Intimacy and Other Short Stories*. Tr. L. Alexander. New York: New Directions Publishing.

—— (1956[1943]) *Being and Nothingness*. Tr. H. Barnes. New York: Washington Square Press.

—— (1964) *The Words*. Tr. B. Frechtman. New York: Random House Publishing.

—— (1976a) *Dirty Hands* in *No Exit and Three Other Plays*. Tr. I. Abel. New York: Random House Publishing.

—— (1976b) *No Exit* in *No Exit and Three Other Plays*. Tr. S. Gilbert. New York: Random House Publishing.

—— (1991). *Transcendence of the Ego*. Tr. F. Williams and R. Kirkpatrick. New York: Hill and Wang Publishing.

—— (1992) *Notebooks for an Ethics*. Tr. D. Pellauer. Chicago: University of Chicago Press.

—— (1993) *What is Literature?* Tr. B. Frechtman. New York: Routledge.

—— (1998) *Existentialism and Human Emotion*. Tr. B. Frechtman and H. Barnes. New York: Kensington Publishing Corporation.
—— (2007) *Existentialism is a Humanism*. Tr. C. Macomber. New Haven: Yale University Press.
Sartre, J.-P. and Benny Levy (1996). *Hope Now: the 1980 Interviews*. Tr. A. van den Hoven. Chicago: University of Chicago Press.
Wang, S. (2009) *Aquinas and Sartre: On Freedom, Personal Identity, and the Possibility of Happiness*. Washington, D.C.: Catholic University Press.
Wike, V. S. (1994) *Kant on Happiness in Ethics*. Albany: State University of New York Press.
Wood, A. W. (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

Part IV
Metaphilosophy

10

Sartre's Method: Philosophical Therapy or Transcendental Argument?

Katherine Morris

There is no doubt that Sartre offers descriptions of everyday experiences and makes phenomenological (indeed, phenomenological-ontological) claims about the essential structures of the lifeworld and what he calls "human reality". The question I want to address here is where Sartre sees his philosophical work to lie in relation to these sorts of offerings.[1] I will consider two rather different types of answer. On the first, what I will call the transcendental reading, where the real philosophical work comes is in demonstrating, via a transcendental argument, that the phenomenological claims are conditions for the possibility of the everyday experiences Sartre describes.[2] On the second, what I will call the therapeutic reading, the phenomenological claims are elicited (in a manner to be indicated) from the descriptions; to be sure, there is serious philosophical work involved in doing this well (although a different kind of philosophical work than the transcendental reading supposes), but there is also important philosophical work to be done, unrecognized by the transcendental reading, in loosening the hold of intellectual 'bad faith' in the grip of which his descriptions (and as a consequence, his phenomenological claims) cannot be acknowledged as accurate.[3] To put

[1] I am using the expression "philosophical work" here to avoid the problematic and contested term "argument".

[2] E.g. Sacks (2005a), Gardner (2011); Gardner finds not only transcendental argumentation but transcendental idealism in Sartre.

[3] See Morris (2008: ch. 2) on "intellectual bad faith". "Intellectual" is by way of contrast with "everyday" (affecting someone's day-to-day conduct); intellectual bad faith affects how philosophers or psychologists reason qua philosophers or psychologists, what possibilities they are open to, what they take for granted, etc.

the point otherwise, whereas transcendental arguments are typically understood to begin from indisputable premises – namely, in the case of the transcendental reading of Sartre, the descriptions of everyday experiences – the therapeutic reading recognizes that these 'premises' are *disputed*; thus on the therapeutic reading, there is work to be done not just in moving from descriptions to phenomenological claims but in breaking down bad-faith-generated resistance to the descriptions.

I will raise some questions about the first reading, but I am less interested in arguing against it than in exhibiting the second as a viable possibility. I will do this in connection with a handful of detailed 'case studies' from *BN*. The first section looks at the move from descriptions of everyday experiences to phenomenological-ontological claims. The second looks more closely at the descriptions with a view to showing why someone in the grip of intellectual bad faith might resist them and how one might see Sartre as endeavouring to break down that resistance. The third considers and rebuts objections that have been made to the therapeutic reading by advocates of the transcendental reading, and suggests an objection to the transcendental reading.

1 Descriptions of experience and phenomenological-ontological claims

Many of Sartre's descriptions of everyday experience are well-known; I offer here a handful of examples. The argument I am urging necessitates my quoting from them at some length.

(1) *Pierre's absence from the café*. "I have an appointment with Pierre at four o'clock. I arrive at the café a quarter of an hour late ... Will he have waited for me? I look at the room, the patrons and I say 'He is not here' ... When I enter this café to search for Pierre, there is formed a synthetic organisation of all the objects in the café, on the ground of which Pierre is given as about to appear ... I am witness to the successive disappearance of all the objects ... in particular of the faces, which detain me for an instant (Could this be Pierre?) and which as quickly decompose precisely because they 'are not' the face of Pierre ... This figure which slips constantly between my look and the solid, real objects of the café ... is Pierre raising himself as nothingness on the ground of the nihilation of the café" (*BN* 9–10).

(2) *The coquette*. "She knows very well the intentions which the man ... cherishes regarding her. She knows also that it will be necessary for her sooner or later to make a decision. But she does not

want to realise the urgency ... she does not want to see possibilities of temporal development which his conduct presents ... The man ... appears to her sincere and respectful as the table is round or square ... In order to satisfy her, there must be a feeling which is addressed ... to her full freedom ... But at the same time the feeling must be wholly desire; that is, it must address itself to her body as object ... But then suppose he takes her hand ... To leave the hand there is to consent in herself to flirt, to engage herself. To withdraw it is to break the troubled and unstable harmony which gives the hour its charm ... We know what happens next; the young woman leaves her hand there, but she *does not notice* that she is leaving it ... because it happens by chance that she is at this moment all intellect" (*BN* 55–6).

(3) *The man at the keyhole*. "Let us imagine that moved by jealousy, curiosity, or vice I have just glued my ear to the door and looked through a keyhole. I am alone and on the level of a non-thetic self-consciousness ... [My attitude] is a pure process of relating the instrument (the keyhole) to the end to be attained (the spectacle to be seen) ...

But all of a sudden I hear footsteps in the hall. Someone is looking at me! What does this mean? It means that I am suddenly affected in my being and that essential modifications appear in my structure ...

First of all, I now exist as *myself* for my unreflective consciousness ... all of a sudden I am conscious of myself as escaping myself ... in that I have my foundation outside myself ... I *am that Ego* ... I discover it in shame and, in other instances, in pride. It is shame or pride which reveals to me the Other's look and myself at the end of that look" (*BN* 259–61).

All of these examples, in one way or another, are aimed at revealing essential structures of human reality (and in some cases these revelations are explicitly signalled by very Kantian-sounding questions).[4] Just prior to (2), for example, Sartre says that he aims to "fix the conditions for the possibility of bad faith" and asks "What must the being of man

[4] Gardner (2011: 59–60) calls attention to the explicitly Kantian-sounding agenda of Sartre's chapter on transcendence; unless one supposes that every way of responding to a Kantian question is a transcendental *argument*, this goes no way toward establishing the case for a transcendental reading of Sartre as I have defined this.

be if he is to be capable of bad faith?" (*BN* 55). The answer is that bad faith "utilizes the double property of the human being, who is at once a *facticity* and a *transcendence*" (*BN* 56), and, more generally, human reality is "a being which is what it is not and is not what it is" (*BN* 58). Example (1) is explicitly aimed at the question of whether a negative judgement "causes non-being to appear at the heart of being or merely limits itself to determining a prior revelation" (*BN* 9); the description is taken to demonstrate the latter: "what is offered to intuition is a flickering of nothingness" which then "serves as foundation for the judgement – 'Pierre is not here' " (*BN* 10). It emerges from the description that nothingness is not "derived from" "being" (i.e. being-in-itself) (*BN* 11), leading to the dominant question of the chapter: "where does nothingness come from?" (*BN* 11). The answer Sartre offers is that there is "a being by which nothingness comes to things" (*BN* 22, in italics in original), namely Man (*BN* 24), and, indeed, this reveals an essential structure of human reality: "Man is the being through whom nothingness comes to the world" (*BN* 24). (Sartre goes on to connect this "possibility which human reality has to secrete a nothingness which isolates it" to freedom (*BN* 24), but we need not pursue this further here.) In the case of example (3), the explicit question addressed is "What does being seen mean for me?" (*BN* 259); ultimately this example contributes to the elucidation of "the nature of the look" (*BN* 280) and (correlatively) "the essential structures of being-for-others" (*BN* 297), as well as to a phenomenological-ontological claim (again expressed in a rather Kantian fashion): "this consciousness [of being an object] can be produced only in and through the existence of the Other" (*BN* 271).

Our question is this: how does Sartre get from the descriptions to the phenomenological-ontological characterization of human reality? We can begin by considering in general terms how those who take the transcendental reading of Sartre would reply to this.

1.1 The transcendental reading of Sartre

There are multiple issues around the nature and defensibility of transcendental arguments in general, and hence a vast literature on this topic, which cannot be reviewed here; for our purposes, it will be most useful to look at a particular interpretation of transcendental arguments by an author who wishes to read Sartre transcendentally, namely Sacks (2005a and 2005b).[5] His central concept is that of "situated thought", whose

[5] This focus is further justified by the fact that Gardner (2011: 52) accepts Sacks's analysis.

content "is informed by the way the subject is epistemically positioned in relation to the facts that make the bare proposition true" (2005b: 446). For instance, if I am situated in front of a tree in the garden, the thought "There is a tree in my garden", if understood unsituatedly, does not license the inference "There is a tree in front of me", but does, he argues, if understood situatedly – and indeed the inference thus understood is, he argues, both synthetic and a priori (2005b: 445).[6] Such inferences from situated thoughts are held to constitute the crucial transcendental premise in transcendental arguments represented as (possibly extended) *modus ponens* arguments.

How do such situated thoughts relate to experiences, according to Sacks? First, experience is articulated (else it would not be experience), but this is not to say that it is linguistically articulable by the subject; "it is not even to say that that articulation is fully cognitive", and the point of talking of 'situated thought' "is precisely to focus on that articulation, and to make it cognitively salient" (Sacks 2005b: 444). From this perspective, what I have been calling 'descriptions of experience' would presumably be expressions of situated thoughts; the move from these to the phenomenological-ontological claims would be a matter of inference of the sort exemplified in the inference to "There is a tree in front of me." Needless to say, the actual transcendental arguments involved will be far more complex than this (and we will look more closely in a moment at Sacks's proposed reconstruction of Sartre's argument in relation to the Look and being-for-others (Sacks 2005a)), but "situated thoughts feature as the grounds for central moves" in such arguments (Sacks 2005b: 451).

There is something deeply attractive about this proposal, but let me raise some questions.

(i) Sacks insists that it is "in the very nature of anything properly called a transcendental argument or proof that it *must* have" a premise which "the sceptic cannot doubt" (2005b: 452, italics his); if the premise in question is supposed to be the situated thought, and if this is supposed to be what is expressed by what I am calling a description of experience, then Sartre's examples do not meet this condition; this is the topic of the next section.

---

[6] One might here ask whether interpretations of Sartre as offering transcendental arguments might be better off utilizing the notion of the "lifeworld *a priori*" (Husserl 1970: III.A §36) as opposed to the Kantian synthetic *a priori*; I cannot pursue this question here.

(ii) I am reluctant to talk of *inference* in Sacks's central case. He himself raises a parallel question in respect of the term 'argument', because, as he rightly notes, "[f]ormal deductive arguments range over, and turn on, relations between propositions or sentences", on which basis he prefers the term 'transcendental proof' for this move within a transcendental argument (2005b: 451). It seems to me that exactly the same point may be made of the term 'inference'. But my hesitation about the use of the term 'inference' is not, I think, mere linguistic squeamishness: Sacks argues that his paradigmatic inference is synthetic on the grounds that "there is more in the thought that there is a tree in front of me than there is in the thought that there is a tree in the garden" (2005b: 445); but is that not just what *isn't* the case if the second thought is understood *situatedly*? It seems to me that Sacks's term 'ampliative' (e.g. 2005b: 443) captures his own move far better than 'inference': what the second thought does is to *amplify*, to unpack, to make explicit what is implicit in the first thought understood situatedly. To be sure, if one were to represent this 'ampliative move' *formally*, it would have to be as an inference; this is presumably what Sacks has in mind when he says that "the material implication at the propositional level merely *shadows*" "the primary ampliative move" (2005b: 443, italics added); the move itself, however, is *not* an inference. (Similar issues dog the Cartesian *cogito*, which Sacks also uses to exemplify his notion of situated thought, 2005b: 446.) Whether we can still talk about 'transcendental *proofs'* once we have done away with the notion of inference is moot; whether we can talk about 'transcendental *arguments'* (of which such transcendental proofs were meant to constitute a central part) will depend in addition on Sartre's text, and we address this question in the next sub-section.

Earlier I used the undefined term 'elicitation' to characterize how the therapeutic reading sees Sartre's move from descriptions to phenomenological-ontological claims. Let me say a bit more about this before we look more closely at what Sartre actually does. I will suggest that the phenomenological ontology is drawn from the descriptions of experience in two steps: first, via the sort of move that Sacks calls 'ampliative'; as long as we are clear that this does not involve an *inference*, his term may help to capture the notion of 'meaning' invoked in such Sartrean questions as 'What does being seen mean for me?' Second, in order to get from the particular to the general, from the meaning of particular experiences to essences and ontological claims which have the status

of necessity, what I am calling 'elicitation' requires us to uncover 'the principle of the series' (*BN* xxii; this is Sartre's definition of 'essence'). The main text of *BN* is largely inexplicit about the procedure for doing so, although Sartre does refer to "an intuition of essences" in the introduction to *BN*, saying that "phenomenal being" "manifests its essence as well as its existence" (*BN* xxii). I will suggest that we are meant to *see* that the particular experiences he describes are typical, and that the meaning 'amplified' from them in the first step is generalizable, necessarily, to all instances of that type. Crucially, this is also not an *inferential* step.[7]

1.2 Sartre's moves from descriptions to phenomenological-ontological claims

Let us, in Wittgenstein's phrase, 'look and see' whether Sartre's text supports the transcendental reading.

(1) *Pierre's absence from the café*. How does Sartre get from this description to the phenomenological-ontological claim that Man is the being by which nothingness comes to things? The description, if accepted, shows that there is an 'intuition' of non-being which is prior to judgement: we experience what Sartre calls *négatités* (*BN* 21). Sartre draws from this that nothingness "haunts" being and "will never be derived from" it; although he develops a different example on *BN* 11, he offers us the materials for this claim in his description of the café, which, "with its patrons, its tables, its booths, its mirrors, its lights, its smoky atmosphere ... is a fullness of being ... Similarly Pierre's actual presence in a place which I do not know is also a plenitude of being" (*BN* 9). Such 'real and objective facts ... and positive events' seem to leave no room for negation (*BN* 11). So far, it seems to me, he is simply drawing out what is implicit in the original description of experience. Can we say the same about his answer to the principal question, namely "where

[7] It has been suggested that on the reading here proposed, "[t]he ontological talk in *Being and Nothingness* would then be either a rhetorical shadow cast by Sartre's map of human phenomenology, or the result of a simple, non-transcendental inference from the appearances" (Gardner 2011: 51). This is a false dilemma. It should be clear by now that I reject the first; the ontology which figures in *BN*'s very subtitle is absolutely central to his philosophical project. But it will also, I trust, be clear that I reject the second. I reserve discussion of the term 'appearances' for §3, but neither 'simple' nor (more importantly) 'inference' applies to the procedure I highlight.

does nothingness come from?" He takes his time about getting to his answer. *BN* I.1.iii considers and rejects Hegel's dialectical concept of nothingness, *BN* I.1.iv considers and rejects Heidegger's phenomenological notion, and then *BN* I.1.v reiterates the question, having sharpened up its urgency: "if Nothingness can be conceived neither outside of Being, nor in terms of Being, and if ... since it is non-being, it can not derive from itself the necessary force to 'nihilate itself,' *where does Nothingness come from*?" (*BN* 22, emphasis in original). He more or less immediately draws the conclusion "that there must exist a Being (this can not be the in-itself) of which the property is to nihilate Nothingness ... a being by which nothingness comes to things" (*BN* 22, last clause in italics in original). This clearly enough *is* an inference, and is more or less set out as a formal argument; we will return to it. The claim that the Being there referred to is Man seems to stem principally from the observation that *négatités* (including Pierre's absence from the café) "derive their origin from an act, an expectation, or a project of the human being" (*BN* 24). What strikes me here is that while there clearly is an *argument* which depends on a 'situated thought' (perhaps expressible as "Pierre is not here"), and which might therefore be called a transcendental argument, it is far from clear that Sartre's phenomenological-ontological claim that man is the being through whom nothingness comes to the world *depends* on it. Sartre fancies himself a scholar and feels the need to review what Hegel and Heidegger have said about nothingness before making his own phenomenological-ontological claim, but it looks as though *BN* 12–23, including the point where he sets out what is more or less a transcendental argument, could simply have been missed out and he would still be in a position to make his phenomenological-ontological claim; these pages are nothing but window-dressing. Simply by amplifying what is implicit in the initial description of Pierre's absence from the café, Sartre can say two things, in addition to the claim that this *négatité* precedes any negative judgement: first, that this *négatité* does not depend on being-in-itself; second, that this *négatité* does depend on my expectation of seeing Pierre. There is, of course, more needed in order to generalize these claims beyond the *particular négatité* so as to be able to say that no *négatités* depend on being-in-itself and that all necessarily depend on human acts, expectations and projects, and thence to identify as an essential structure of human reality its possibility of secreting nothingness. Sartre claims – having considered a series

of negative judgements made "to amuse myself" ("Wellington is not in this café, Paul Valéry is no longer here, etc." (*BN* 10)) – that "[t]his example is sufficient to show that non-being does not come to things by a negative judgement" (*BN* 11) (although he also considers numerous other examples of *négatités*). In order to make sense of this claim, we may say, as I suggested earlier, that we are meant to *see* that everything which Sartre says about the meaning of this particular example is equally and necessarily true of every *négatité*. A full account of the means of arrival at the claim that the possibility of secreting nothingness is an essential structure of human reality might take something like the same form, i.e. we are meant to *see* that it is not just Sartre but all of us, necessarily, who undertake acts and who have expectations and projects.

(2) *The coquette*. Again, how does Sartre actually get from passage (2) to the phenomenological-ontological claim? Well, first he takes note of the coquette's use of "various procedures in order to maintain herself in bad faith": disarming her companion's actions "by reducing them to being only what they are" while at the same time permitting herself "to enjoy his desire" and hence, insofar as she apprehends it as "not being what it is", recognizing its transcendence; and "while sensing profoundly the presence of her own body", she "realizes herself as *not being* her own body", treating it as "a passive object to which events can *happen*" (*BN* 56). This much could easily be regarded simply as an amplification of the situation described in (2), one which highlights the notions of facticity and transcendence. These observations lead him to a general characterization of bad faith as "a certain art of forming contradictory concepts", and he then claims that these contradictory concepts utilize "the double property of the human being, who is at once a *facticity* and a *transcendence*" (*BN* 56) – i.e. he identifies the relevant essential structure of human reality *here*, with no further ado. He goes on to call attention (*BN* 58) to further 'duplicities' in addition to facticity and transcendence (being-for-itself/being-for-others, being-in-the-world/being-in-the-midst-of-the-world, etc.), the second in particular also visible in the case of the coquette, and thence arrives at his more general formulation about human reality. In this case there is nothing that even looks like a transcendental argument here. We have, once again, some amplification of the original description which presents *this instance* of bad faith as playing with facticity and transcendence; we have a generalization (bad faith is "a certain art of forming contradictory concepts")

which suggests that *all* instances of bad faith, necessarily, similarly play with facticity and transcendence or perhaps the other dualities, and we have a phenomenological-ontological claim about the essential structure of human reality (human beings are both facticity and transcendence). Again, one would like a more explicit account of the route to the generalization and to the phenomenological-ontological claim, but we can say much what we said about the previous case.

(3) *The man at the keyhole*. In this case, we have an extant reconstruction (Sacks 2005a) of Sartre's supposed transcendental argument with which to engage. Sacks identifies two important movements of thought in Sartre's discussion: the first (which he takes to be successful) involves showing that the Look is a necessary condition of my awareness of myself as an object; the second (which he deems unsuccessful) is "the inference to reality": "the worry is that the prerequisites of coming to know what I am can be met without actually going so far as to imply that there are others" (Sacks 2005a: 291). Regarding Sacks's discussion of the first movement of thought, I will suggest that Sacks's discussion loses important features of Sartre's phenomenology of the Look. As for the second, I will argue (i) that we have once again to do with a move which does not involve an inference, and this is because (ii) the existence of the Other is elicited (not inferred) from Sartre's phenomenology of the Look once that *is* fully developed.

Sacks argues that "what the Look yields that is unique to it … can be gauged from my response to it": the shame that immediately overwhelms me, such that "I know that someone else realizes what I was doing" with "no margin of error", shows that "what they have caught at it is not my body, but me" (2005a: 287). Moreover, "the only thing that could give rise to an awareness that I was inalienably an object, would be … my coming to perceive that I – this very subjective consciousness that I am for myself – have been captured as an object" (2005a: 288). Finally, "in being observed I realize that … the logically adequate methods for other-ascription of these predicates [namely, person-predicates or 'P-predicates' in Strawson's terminology, here referenced by Sacks] are the ones appropriately being employed by the Other whose Look is directed at me" (2005a: 289). This is useful, but it seems to me to miss certain crucial features of the phenomenology of the Look as Sartre develops it. First, we should be reluctant – given Sartre's characterization of human reality as "being what it is not and not being what it is" and his consequent claim that

"the principle of identity" does not apply to the For-itself (*BN* 58) – to accept an interpretation which speaks of the "numerical identity of experiencing subject and perceived object" (Sacks 2005a: 287). Yes, the shame that I feel in the scenario described "is the *recognition* of the fact that I am indeed that object which the Other is looking at and judging" (*BN* 261); but clearly enough, I am that object "in the mode of *being what I am not*" (cf. *BN* 60). Second, and following on from this, Sacks's interpretation misses the fact I am not the foundation of this object which I am: it is not even "the indirect, strict effect of my acts as when my shadow on the ground or my reflection in the mirror is moved in correlation with the gestures which I make. This being which I am preserves a certain indetermination, a certain unpredictability" (*BN* 261–2). Would this be the case if all that the look revealed was that another was applying P-predicates to me? The central point is surely that "the Other as a look" is "my transcendence transcended" (*BN* 263). The Other's look is at "the very center of my act as the solidification and alienation of my own possibilities" as well as "the alienation of the world which I organize" (*BN* 264). That is, the Other is not simply applying P-predicates (on "logically adequate criteria") to me but judging me and responding to me as an object in *his* world, and that world is structured by *his* freely chosen projects and values. "Thus being-seen constitutes me as a defenceless being for a freedom which is not my freedom" (*BN* 267; cf.: "by fixing my possibilities the Other reveals to me the impossibility of my being an object except for another freedom", *BN* 270).

Once it is acknowledged that the look reveals a freedom which is not mine, we are in a position to make the phenomenological-ontological claim: to recognize "the indubitable existence of this Other for whom we are" (*BN* 282). Our resistance to solipsism "is based on the fact that the Other is given to me as a concrete evident presence which I can in no way derive from myself and which can in no way be placed in doubt nor made the object of a phenomenological reduction" (*BN* 271).[8] The phenomenological-ontological claim is elicited from the description in much the same manner as the previous two cases: there is no transcendental argument here.

One final point: it seems to me that those who take the transcendental reading of Sartre must on these grounds find his text *wanting*: his

[8] The chapter of course continues at length, partly replying to objections, partly clarifying more fully "the fundamental relation of the Me to the Other" (*BN* 282), partly clarifying what the Other thus revealed is, etc. But the fundamentals are there, I submit, by this stage in the chapter.

arguments need to be reconstructed in order to be visible.[9] It may be felt that exactly the same charge can be levelled at me: I have acknowledged that Sartre is not very explicit about the processes which mediate between the meanings (as amplified from the descriptions) and the phenomenological-ontological claims; his claim in the introduction that "phenomenal being" "manifests its essence as well as its existence" is as explicit as he gets. But the point about the therapeutic reading is that it sees the centre of gravity of Sartre's task as lying elsewhere: if he can persuade his interlocutors to accept the descriptions, there is no particular difficulty about getting them to accept the phenomenological-ontological claims. The really tricky part, according to this reading, lies in persuading his interlocutors to accept the descriptions, and this is because the interlocutors he is addressing are themselves in the grip of intellectual bad faith.

2 Bad-faith resistance to these descriptions

We noted earlier that the transcendental reading requires the 'premise' of the argument to be 'indisputable'. To expand upon my objections to this, it presupposes that there will be no resistance to the claim that experience is articulated (cf. Sacks 2005b: 444), and likewise seemingly presupposes that there will be no resistance to the way in which the situated thought (which makes explicit the articulation of experience) is expressed in language. Both are false; the claim that experience is articulated is widely resisted by philosophers of an empiricist stripe (notwithstanding Kant's arguments). To establish that the second presupposition is false, I will call attention to the numerous ways in which Sartre's descriptions discussed earlier are apt to be anything but undisputed. (I confine myself to the first two, since the third moves almost immediately into "the meaning of being seen" and hence into amplification – not that the line between these two is always clear.)

(1) *Pierre's absence from the café*.
   – "*When I enter this café to search for Pierre, there is formed a synthetic organisation of all the objects in the café, on the ground of which Pierre is given as about to appear*"; "*the faces … detain me for an*

[9] Baker has made a similar point about those who want to read Wittgenstein's remarks in *PI* §§243–315 as a *reductio ad absurdum* (Baker's own reading of Wittgenstein is therapeutic through and through): "if one considers the private language argument as a *reductio* of Cartesian dualism, it manifests a large number of defects" (2004: 116).

instant … and … quickly decompose precisely because they 'are not' the face of Pierre". These parts of the description presuppose the Gestalt psychology notion that experience is structured into figure and ground; according to Köhler, the figure is given as "solid" and "substantial", while the ground appears comparatively "empty" or "undifferentiated" (1947: 120). Köhler was explicitly arguing against empiricist psychology, and of course Merleau-Ponty (in *PP*) put his anti-empiricist arguments to work in an even more thoroughgoing way. Thus, surely, these parts of the description will be resisted by empiricists.

– "*This figure which slips constantly between my look and the solid, real objects of the café … is Pierre raising himself as nothingness on the ground of the nihilation of the café*". The main point of this description is to reveal an intuition of an absence, of a *négatité*; *any* philosopher (a determinist, a positivist, a realist) who supposes that all we encounter in the world are positive existences will resist this.

2) *The coquette*.

– "*[S]he does not want to see possibilities of temporal development which his conduct presents*": a behaviourist would refuse to accept that behaviour ("conduct") could "present" "possibilities of temporal development".

– "*The man … appears to her sincere and respectful as the table is round or square*": those who take character traits to be determinate properties of human beings will not see the hint of bad faith in this description.

– "*the feeling must … address itself to her body as object*": those who take the body to be nothing but an object will find the last two words peculiar and redundant.

The difficulties for passage (2) are compounded if we include in the description the coquette's use of "various procedures in order to maintain herself in bad faith" (*BN* 56):

– first, she "disarms" her companion's actions "*by reducing them to being only what they are*" while at the same time permitting herself "to enjoy his desire" and hence, insofar as she apprehends it as "*not being what it is*", recognizing its transcendence: anyone who supposes that the principle of identity applies to human reality (again, this includes empiricists, realists, positivists …) will certainly resist this;

– second, "while sensing profoundly the presence of her own body", she "*realizes herself as* ***not being*** *her own body*", treating it as "*a*

> *passive object to which events can* ***happen****"*; once again, those who take the body to be nothing but an object will find no trace of bad faith in either of the last two clauses.

Thus Sartre's descriptions *cannot* be understood as playing the role of indisputable 'premises'.

The next point is that all of the philosophical positions which would lead to resisting these descriptions can be regarded as forms of intellectual bad faith; indeed, they all, in one way or another, regard either human beings or the lifeworld or both as *things* (as positive existences, as things which "are what they are and are not what they are not", as not pointing beyond themselves, as not having significance, etc.). (Merleau-Ponty (in *PP*) would speak of the 'prejudice of objective thought' where I speak of 'intellectual bad faith'.) Now, I don't claim to be able to demonstrate that Sartre so regarded them, but he certainly does occasionally charge *some* of his philosophical interlocutors with intellectual bad faith (*BN* 113, 249),[10] and it would be surprising if he were unwilling to level the charge more widely where appropriate, especially given his views about the prevalence of everyday bad faith.

Of course, even if Sartre did see these sources of resistance as exhibiting a kind of bad faith, it clearly doesn't follow that he saw any part of his philosophical project to try to free his interlocutors *from* that bad faith. Again, I don't claim to be able to demonstrate that he did. All I want to do is to open up this possibility and to ask what his philosophical procedures might look like if seen in this light. What would it be for a book to engage in a therapeutic process?

Well, we have acknowledged examples of such books, most notably Wittgenstein's *PI*.[11] Although Wittgenstein does not use the expression 'bad faith', he does use expressions which carry something of the same import ('prejudice', 'dogma', 'superstition' etc.) and he explicitly sees

---

[10] Note also his claim that "[p]sychological determinism, before being a theoretical conception, is first of all an attitude of excuse" (*BN* 40), which, without actually using the term 'bad faith', clearly implies it.

[11] It may be thought that there is something perverse about my reading Sartre in a way that makes him closer to Wittgenstein (of whom Sartre said that he "would rather read 'thrillers' ", 1967[1964]: 49) than to Kant (an acknowledged predecessor). I chose Wittgenstein as a comparator solely because there are well-known readings of him as a therapeutic philosopher. In fact, I see this reading as bringing Sartre closer to Merleau-Ponty (and possibly even to Heidegger) than to Kant – a far less perverse claim – but these two phenomenologists are less widely recognized as therapeutic philosophers.

his philosophical task in *PI* as 'therapeutic'. Wittgenstein recognizes, as Sartre surely does, that there can be no *recipe* for ridding someone of intellectual prejudices or intellectual bad faith, any more than there could be such a thing for ridding someone of everyday prejudices or everyday bad faith; the problem is exacerbated when (as in these cases) the intellectual prejudices or bad faith in question are embedded in the culture. But we can see some of Wittgenstein's strategies for attempting this task in *PI*; the question will be whether we can see similar strategies in *BN*, if we look at *BN* in this light.

The crucial point about everyday bad faith is that it "does not hold the norms and criteria of truth as they are accepted by the critical thought of good faith" (*BN* 68); intellectual bad faith, too, is characterized by what might be called a 'perverse' attitude toward evidence and logical argument. As Wittgenstein puts it: "dogma [prejudice, intellectual bad faith] is expressed in the form of an assertion, and is unshakable, but at the same time any practical opinion *can* be made to harmonize with it; admittedly more easily in some cases than in others…This is how dogma becomes irrefutable and beyond the reach of attack" (1980: 28). Thus the therapeutic task will require strategies other than the provision of evidence and logical argument. The task itself is multifaceted: one must get one's interlocutor to *recognize* that what they want to say is at odds with the way we experience ourselves and the lifeworld; one must make the interlocutor *uneasy* about this fact, to generate inner conflict ("'But this isn't how it is!' 'Yet this is how it *must* be!'", cf. *PI* 112); and one must offer him a *way out* of this inner conflict, attempting to detach him from the picture which holds him captive (here, a picture of human beings and the lifeworld as 'things', although this picture takes many forms) by making it clear that there is no 'must' about their picture, that it really is possible to accept that our ordinary experience reveals how things are.[12]

In Wittgenstein's case, many of the features of *PI* which strike philosophical readers as peculiar (and are liable to be dismissed as mere stylistic quirks) make perfect sense once we see him as engaging in one or another of these therapeutic tasks: e.g. his frequent use of questions, metaphors, analogies, humour, irony, peculiarities of punctuation, etc. We see many of these 'stylistic quirks' in Sartre's writing too, as well as some peculiar to him: his paradoxical modes of expression, his wordplay, his penchant for negative-sounding language, his hyperbole. These

[12] Cf. Morris (2007).

too, I submit, make sense once we see him as engaging in one of these therapeutic tasks.

It would be an article in-itself to make this case in detail; I will just sketch a handful of examples here. First, he may be seen as using *humour* to a number of therapeutic ends. One might be to charm his readers into a willingness to embrace a phenomenologically accurate characterization of human reality ("thanks to transcendence, I am not subject to all that I am … I leave my tattered garment in the hands of the fault-finder" (*BN* 57)); even his penchant for negative-sounding language (e.g. his focus on shame and vulgarity in his discussion of being-for-others) may be seen as a kind of 'charm offensive', though not one that works with everyone. A second use of humour can be understood in the light of his own proposal that play and irony are activities which stand in opposition to what he calls "the spirit of seriousness", a "dismissal of human reality in favour of the world" (*BN* 580), in which man "takes himself for an object" and in which, in particular, he sees values as written into the world independently of his own free choices; this "spirit" is a manifestation of bad faith and is equally seen in intellectual bad faith, and play opposes it because it "releases subjectivity" and opens up possibilities (*BN* 580–1). Thus he pokes fun at bad-faith conceptions; this is most obvious in his critique of the everyday bad faith of the anti-Semite, who, for example, admits that Jews are intelligent because that makes them more dangerous, and allows him to disdain intelligence because it is Jewish: "The true Frenchman … does not *need* intelligence" (1948[1946]: 25), but there are examples of its use against intellectual bad faith, as when he characterizes Bergson's conception of the being of the past as "preserving for it the existence of a household god" (*BN* 109–10). Second, his paradoxical modes of expression (paradigmatically, human reality as "a being which is what it is not and is not what it is" (*BN* 58)) may also be seen as playful; alternatively, they may "have been conceived in this form explicitly to shock the mind and discountenance it by an enigma" (as he says of such everyday expressions as "He has become what he was" (*BN* 57)), in which case they serve an equally therapeutic purpose of forcing his readers either to dismiss Sartre altogether or to reflect on their own ontological commitments and, just possibly, becoming uneasy about them. Third, Sartre's well-known hyperbole may be seen specifically as a therapeutic measure designed to caricature, but thereby to highlight the essential features of, everyday experience so as to re-orient the thinking of his interlocutor. For example, his announcement that "[m]y body is co-extensive with the world" (*BN* 318) provides a counterweight to the picture according to which the body is simply

one more physical object in the midst of the world. And when he makes the – both hyperbolic and negative-sounding – claim that when another person comes along, "suddenly an object has appeared which has stolen the world from me" (*BN* 255), this is meant to shift our thinking off from the idea that what we see when we see another human being is simply a physical object that moves.

There is a great deal more to be done to make the case here wholly persuasive, but I hope to have shown that *BN can* be read in this light.

3 Objections and replies

I end by considering objections which have been made to the therapeutic reading here proposed, and an objection to the transcendental reading.

First, the therapeutic reading has been represented as follows: "if the phenomenological descriptions he [Sartre] gives are convincing, then arguments are not needed" (Gardner 2011: 51). I take this to be a criticism of the therapeutic reading since it seems to place Sartre's procedure outside the bounds of philosophy, which surely centrally involves *arguments*. I would make two points here (apart from the obvious one: what counts as an argument?): first, that the question must be: convincing *to whom*? The centre of gravity of the therapeutic reading lies in the recognition that many thinkers have motives – sc. what I am calling intellectual bad faith – for *refusing* to find his descriptions convincing. Second, the criticism conflates two points at which arguments might be thought to be needed. (1) It seems to me that there is a clear sense in which arguments are not needed for the transition from description to phenomenological-ontological claims, because the latter are (in a sense to be clarified below) things which we always *already* 'know' to be true;[13] this does not entail that there no philosophical work going on here, since much philosophical skill is needed to elicit (in the sense outlined in Section 1) the phenomenological-ontological claims from the descriptions, and, indeed, much philosophical skill is needed in describing particular experiences in such a way that they sustain the move from particular to general. (2) It might be thought that arguments are needed

[13] This is a point which Sartre makes about the Cartesian *cogito* as well: it likewise is not an argument. And this passage: "Similarly my resistance to solipsism – which is as lively as any I should offer to an attempt to doubt the *cogito* – proves that I have always known [possessed certainty: see below] that the Other existed" (*BN:* 251).

to get recalcitrant interlocutors to accept the descriptions of everyday experience; but here the point is that 'argument' as we would ordinarily understand it is useless – by definition. No one was ever 'cured' of bad faith (be it everyday or intellectual) by 'sheer force of argument'; as noted earlier, it is of the essence of bad faith that it has what may be called a 'perverse' attitude to evidence and argument (cf. *BN* 68).

Second, it has been suggested that the therapeutic reading involves viewing Sartre as "turning his back on epistemology" (Gardner 2011: 52). To be sure, I have suggested that Sartre's primary therapeutic targets are (not sceptics but) holders of (bad-faith) ontological positions, in particular those who see the world and human beings through the lens of 'the prejudice of objective thought'; bad faith, whether everyday or intellectual, is in the first instance a (wilful) misconstruction of *ontology*. It is precisely this that interferes with their accepting descriptions that, if accepted, would allow the elicitation of a (non-bad-faith) phenomenological ontology. (This is already a reason for resisting the transcendental reading, given that transcendental arguments are normally taken to be *anti-sceptical*.[14]) But this hardly entails seeing Sartre as "turning his back on epistemology". On the contrary, assuming that 'certainty' (what is sometimes called 'lived certainty') is an epistemological notion, epistemology is central both to his ontological and to his therapeutic project. (His rejection of the "primacy of knowledge" (cf. Gardner 2011: 51) is in part an affirmation of the 'primacy of *certainty*'. It is simultaneously an affirmation of the primacy of 'being' over 'knowing'. At the level of lived certainty, epistemology and ontology in a certain sense come together.) What makes bad faith *bad faith*, as opposed to (say) mistake or confusion, is that we actually 'know better': but the word 'know' is (precisely) out of place. In virtue of being human beings who are in the world, we *live the certainty* that there are *négatités* to be met with in the world, that we are ourselves both facticity and transcendence, that others exist, and so on.

This, however, brings us to one of the most important points of disagreement between the transcendental and the therapeutic readings, and what may be seen as a criticism of the former. Advocates of the

[14] That Sacks (2005a, 2005b) so takes them is manifest. To be sure, bad-faith ontologies can lead willy-nilly to certain forms of scepticism; hence Sartre's 'realist' and 'idealist' founder on "the reef of solipsism" (*BN:* III.1.ii) because they end up unable to *justify* the claim that others exist. As Sartre argues, however, they so desperately do not want to be in that position that they inadvertently slide to their opposites so as to avoid it; and it is their *ontologies* which make a justification of the existence of others impossible.

transcendental reading begin from the supposition that there is a gap to be bridged between 'appearance' and 'reality'. Gardner speaks of "inferences from the appearances" (2011: 51); and Sacks ends up characterizing Sartre as being unable to "make good the inference to reality", despite having done something of great value in identifying "transcendental necessities in the way we construe the empirical world" (2005a: 296). Nothing could more starkly reveal the difference in starting point between Kant and the phenomenologists (at least the existential phenomenologists). Transcendental readers can hardly be unaware that Sartre's literal starting point is the "reduction" of "the existent to the series of appearances which manifest it" and the rejection of a large handful of "dualisms", including of course "the dualism of being and appearance" (*BN:* xxi). How, then, can they suppose that Sartre is trying (and perhaps failing) to make an inference from appearance to reality? Perhaps it is felt that he can't just do away with such dualisms, but where does the burden of proof lie? Kant famously called it "a scandal of philosophy and human reason in general" that there is still no cogent proof for the existence of things outside us (quoted in Heidegger 1927[1962]*:* 203), and proposed his own proof. Heidegger offered the rejoinder that the true "scandal of philosophy" "is not that this proof has yet to be given, but that *such proofs are expected and attempted again and again*" (1927[1962]: 205), and my reading of Sartre sees him as taking up this tradition.[15]

What I have tried to do here is to make out a case for the *possibility* of reading Sartre in a way that does not involve his offering transcendental arguments. I have suggested, in the first place, that notwithstanding Sartre's sometimes explicitly Kantian framing of the discussion, careful attention to the text provides less evidence for any form of transcendental *argument* than for a dual move which I called 'elicitation', involving, first, a (non-inferential) 'ampliative' move from description of experience to 'meaning', and second, a (non-inferential) revelation of 'the principle of the series', i.e. the essence of the phenomenon described and thus part of a phenomenological ontology. I have suggested, in the second place, that far from Sartre's descriptions of experience being 'indisputable' even to the sceptic, they are (1) disputed, but (2) the interlocutors who dispute these descriptions are not sceptics but thinkers in the grip

[15] According to Gardner, Sartre's response to the sceptic, unlike Heidegger's, "incorporates a recognition that sceptical doubts are *meaningful"* (2011: 52). Needless to say, I disagree (and not just because I don't see his engagement as primarily with the sceptic), but it is beyond the scope of the present essay to spell out why; it would involve showing how he can be read as 'dissolving' rather than 'solving' philosophical problems.

of intellectual bad faith, i.e. thinkers with a bad-faith *ontology*. It is the further suggestion that it is possible to read *BN* as aimed at *freeing* these interlocutors from the grip of intellectual bad faith that warrants the term 'therapeutic'.

References

Abbreviations used:
BN: Sartre, J.-P. (1986[1943]) *Being and Nothingness*. Tr. H. E. Barnes. London: Routledge.
PI: Wittgenstein, L. *Philosophical Investigations* (1958). Ed. G. E. M. Anscombe and R. Rhees. Tr. G. E. M. Anscombe. Oxford: Blackwell.
PP: Merleau-Ponty, M. (2011[1945]). *Phenomenology of Perception*. Tr. D. Landes. London: Routledge.
Baker, G. P. (2004). *Wittgenstein's Method: Neglected Aspects*. Ed. K. J. Morris. Malden, Oxford and Carlton: Blackwell.
Gardner, S. (2011). "The Transcendental Dimension of Sartre's Philosophy", in *Reading Sartre*. Ed. J. Webber. Abingdon and New York: Routledge.
Heidegger, M. (1927[1962]). *Being and Time*. Tr. J. Macquarrie and E. Robinson. Oxford: Blackwell. (References to this work are to Heidegger's page numbers.)
Husserl, E. (1970). *Crisis of European Sciences and Transcendental Phenomenology*. Evanston: Northwestern University Press.
Köhler, W. (1947). *Gestalt Psychology*. New York and London: Liveright.
Morris, K. (2007). "Wittgenstein's Method: Ridding People of Intellectual Prejudices", in *Reading Wittgenstein*. Ed. E. Kanterian et al. Malden, Oxford and Carlton: Blackwell.
—— (2008). *Sartre*. Oxford: Blackwell.
Sacks, M. (2005a). "Sartre, Strawson and Others", *Inquiry* 48: 275–99.
—— (2005b). "The Nature of Transcendental Arguments", *International Journal of Philosophical Studies,* 13: 439–60.
Sartre, J.-P. (1948[1946]). *Anti-Semite and Jew.* Tr. G. J. Becker. New York: Schocken Books.
—— (1967[1964]). *Words*. Tr. I. Clephane. Harmondsworth: Penguin.
Wittgenstein, L. (1980). *Culture and Value*. Ed. G. H. Von Wright and H. Nyman. Tr. P. Winch. Oxford: Blackwell.

11

The Transcendental Idealisms of Kant and Sartre

Richard E. Aquila

1 Idealism and phenomenalism

Sartre opens "The Pursuit of Being", his Introduction to *Being and Nothingness*, noting that "considerable progress" can be found in a recent development of "modern thought". What is in question is a new manner – presumably, new at least since Berkeley – of "reducing the existent to the series of appearances which manifest it". Sartre makes it clear that further refinement is in order. But the new approach has at least the advantage of replacing, or promising to replace, some objectionable dualisms with a "monism of the phenomenon" (3/11).[1]

It may be insufficiently appreciated that what Sartre has in mind is a particular approach, not simply to the relation between existents and appearances or (to the extent that there is a difference) phenomena, but to the very *idea* of a "phenomenon". This may be obscured by Hazel Barnes's translation ("The Phenomenon") of the title of the first section: "*L'Idée de Phénomène*". Immediately following initial comment on the issue at hand, Sartre in fact says: "Thus we arrive at the idea of a *phenomenon* (*l'idée de* phénomène) such as we can find, for example, in the

[1] For "appearances", Sartre sometimes says *apparitions*, sometimes *apparences*, generally with no evident difference in intention. However, '*apparition*' can also refer to the "appearing" of an appearance, and we will take note later of an issue turning on this distinction. See footnote 33. Throughout, for English rendering of *L'être et le néant* (Sartre 1943), I follow, with occasional modification, the translation of Hazel E. Barnes (Sartre 1966[1943]). For the French I cite from (Sartre 1976[1943]). Pagination of the form 'xx/yy' displays the reference to the Barnes translation first, and then to the French edition.

'phenomenology' of Husserl or of Heidegger" (4/12). I take the position in this chapter that Sartre not only espouses a kind of "phenomenalism" to which the idea in question is key but – despite his own reading of Kant and his aim to transcend the distinction between realism and idealism (26/30) – is sharing with Kant a view not unreasonably called "transcendental idealism". Of course, this will require avoiding what Sartre himself calls phenomenalism (*phénoménisme*): a view which in Husserl he claims "every moment borders on Kantian idealism" (119/109).[2] In any case, it should become clear why I call the variety of phenomenalism that I attribute to both Sartre and Kant "transcendental phenomenalism", and also why I regard it as a particular form of, but arguably not strictly entailed by, the "transcendental idealism" that I likewise ascribe to them.

As for what Sartre sees himself as sharing with Husserl despite the latter's shortcomings, it seems reasonable to locate this in agreement that, in at least some sense, an object is a "connected (*liée*) totality" of appearances (17/23), and even that objects can in some sense be identified with the "principle" (*raison*) of such totalities (6/13). What he in any case says is that the phenomenalists have "justifiably (*à juste titre*) reduced the object to the connected series of its appearances" (21/26). Assuming that, at least in this context, Sartre means by a "phenomenon" simply an ordinary existent – an ordinary existent object (or "thing") – and being careful to avoid, as he proceeds to emphasize, an objectionably "phenomenalist" view of the *being* of the appearances in question, it may also seem reasonable to take the following to express Sartre's own view:

> Let us understand indeed that our theory of the phenomenon has replaced the *reality* of the thing by the *objectivity* of the phenomenon ... We shall interpret this by saying that the series of [the cup's]

---

[2] Thus with respect to the label "phenomenalism," I of course depart from the terminological sense of many, including Sartre himself. Cf. (Jeanson 1965[1980]: 109): " 'Relative to' can mean either 'absolutely dependent' on or 'always compelling the recognition of .' Phenomenalism understands the relation of being to consciousness in the first [unSartrean] sense." As for Husserl, Sartre alleges that he "has shut himself up inside the *cogito* and deserves – in spite of his denial – to be called a phenomenalist (*phénoméniste*) rather than a phenomenologist. His phenomenalism (*phénoménisme*) at every moment borders on Kantian idealism" (119/109).

> appearances is connected (*liée*) by a principle which does not depend on my whim. (5/13)[3]

It may seem that, in order to attribute this view to Sartre, we simply need to accommodate his insistence that, just as such, each appearance is something *in-itself*. Certainly, the main thrust of his critique of an objectionably phenomenalistic view lies in his claim that, in some sense, each of them "has its own being", "is not supported by any existent different from itself" (7/14), "lay[s] claim (*réclamaient*) to a being which is no longer itself appearance" (17/23), "is already in itself alone (*à elle toute seule*) a *transcendent being*" (23/27), and that "there is a being of the thing perceived – *as perceived* (*un être de la chose perçue en tant qu'elle est perçue*)" (18/24). But it remains to be seen what this all amounts to. On the other hand, once matters have been so broadly construed as to include Husserl as fellow traveller, despite his shortcoming precisely with respect to this objection, one might wonder why Sartre does not include Berkeley as well. The reading that I will offer explains this in what seems to me a particularly cogent way.

Sartre makes clear his reason for excluding Kant. The sort of dualism which the new "monism of the phenomenon" is supposed to overcome is, on his reading of Kant, one that would have us distinguish between phenomenal objects and "things in themselves", or noumena, hidden behind the phenomenal scene. More specifically, a phenomenon for Kant always "point[s] over its shoulder to a true being" (4/12), it "refers (*renvoie*) to the noumenon…behind the appearance" (6/14; perhaps better: "refers *us* to"), and indeed Sartre takes this to be part of the very *idea* of "the phenomenon" for Kant. Obviously, that is a disputed question. But again, what of Berkeley, who is later introduced with more specific respect, not to the question of modern phenomenalism as such, at least as so far characterized, but rather to the question of the

---

[3] For *liée*, Barnes says "bound" here, in the passage quoted just above "united". Cf. Gardner (2009: 39):

> [R]eal, objective existence reveals itself, not of course in any finite sum of appearances, but in a particular *mode* of appearing, one where each individual appearance of an object…refers us to an indefinite number of other possible appearances of that object, according to some 'law' or 'principle' which makes the series of appearances non-arbitrary.…Sartre's attitude is that [this] is already, thanks to Husserl and Heidegger, well-established.

As for Gardner's view of the status of appearances *as such*, see footnote 21.

ontological status of the *appearances* to which phenomena/objects are supposed to be reduced? Certainly, he is guilty of none of what is here charged against Kant. On the other hand, as standardly noted, Berkeley is of course guilty of reducing the "being of appearances" as such to their being *perceived*.[4] But that, whether fairly or not,[5] is a charge that Sartre also levels against Husserl: he has reduced appearances to mere "subjective plenitude[s], the manner in which the subject is affected" (5–6/13; cf. 22/27).

Here I will anticipate the general direction of my approach. Consider any totality or series of appearances to which a given phenomenon is, in whatever sense is supposed to be in question, "reducible", or of which it is supposed to be the relevantly connected totality or series. There is one thing that all three of Kant, Husserl and Sartre want to be able to say (although I won't argue for this with respect to Husserl), and that is completely foreign to Berkeley. All three want to be able to say that, while each appearance within such a series is indeed, from one point of view, a mere appearance *of* the phenomenon in question, it is at the same time, from another point of view, precisely *itself* that phenomenon. Thus:

> *Transcendental Phenomenalism*. Any judgement of an appearance *A*, to the effect that it is a phenomenon (a real existent), is a judgement to the effect that *A* is, from one point of view, a phenomenon and also, from another point of view, an appearance within an infinite series of appearances *of* the very phenomenon in question.

As we might therefore also note: from the former point of view, and precisely given the possibility of the latter, it is also an *at least possible* phenomenal existent in its own right.

This might seem to require regarding the relation between Sartrean phenomena and the appearances to which they are reducible as in effect the same as that proposed for the distinction between "things in themselves" and appearances in Kant, when Kant is interpreted as holding a "two-aspect" as opposed to "two-worlds" view of the

---

[4] We might of course suppose that Sartre has in mind the idea that even "real existents", and not simply the "appearances" to which they are reducible, exist for Berkeley only as perceived, namely, at least by God. But that does not seem in fact to enter into his references to Berkeley.

[5] Cf. Gardner (2009: 44): "Sartre ascribes the claim that the being of the appearance is its appearing to Berkeley and (tendentiously) to Husserl."

distinction.[6] Therefore it might seem to ask us to regard Sartrean phenomena as comparable, at least in that respect, to Kantian things-in-themselves. Kant himself of course distinguishes between the notion of a thing-in-itself in the "transcendental" sense – according to which no phenomenal existent is as such a thing-in-itself, but at most an appearance – and an empirical notion according to which we can distinguish real *phenomenal* existents, as "things in themselves", from appearances and possible appearances *of them* (A45–6/B62–3).[7] The two-aspect reading of Kant, at least in its familiar form, is a reading of the transcendental distinction. In its terms, those phenomenal existents which are, from one point of view, mere appearances are from another point of view not at all distinct from certain things that are (in the transcendental sense) something "in themselves", *of which* they are appearances. But what is in question for us here concerns the empirical not the transcendental distinction, that is, it concerns precisely the relation between phenomenal existents and appearances (or possible appearances) *of them* as such.

On the proposal to be put forth (but with a qualification to be postponed for now), those Sartrean appearances to which phenomenal existents would be reducible are, just as such, something "in themselves", or something with their "own being", *only in the sense* that they are not only indeed, from one point of view, appearances to which phenomenal existents might be reducible but also, from another point of view, precisely themselves at least *possible* phenomenal existents. The alternative, after all, would make no sense. If appearances as such had their own being in any stronger sense – and in particular, precisely in the sense of what Sartre himself calls "being-in-itself" – then that very fact would exclude the sort of ontological multivalence that I am proposing with respect to Sartrean appearances.

This very suggestion may seem to put me in a bind. For without some fuller ontological status, how can there be anything in the first place available *for* determination either as (at least possible) phenomenal existents or (at least possible) appearances thereof? I will speak to this point later. In addition, I will try to show that the consideration

---

[6] As will become clear in the Conclusion, I remain in the present chapter open regarding the possibility of conjoining a "two-aspect" view of the relation between "things in themselves" in the transcendental and the empirical senses with a certain sort of "phenomenalistic" view of the latter. This is a departure from the position I had maintained in Aquila (1983: 88ff).

[7] I cite in standard A/B pagination to the *Critique of Pure Reason*. Unless otherwise noted, translations are those of Guyer and Wood (Kant 1998[1781/1787]).

allowing us to deflect the apparent difficulty also facilitates articulation of a view according to which *phenomena* (i.e. phenomenal existents as opposed to appearances as such) are to be regarded as *both* in some sense "reducible" to appearances and also as having Sartrean being-in-itself. (If so, this would of course entail that appearances or possible appearances *of* phenomena are also to be regarded, from another point of view, as at least *possibly* something with Sartrean being-in-itself.) This would then amount to a kind of *double* two-aspect view of things, combining a certain sort of two-aspect approach to the empirical distinction with a certain sort of two-aspect approach to a "transcendental" distinction between appearances and that which appears through them. It remains dubious to my mind that this is Kant's view, though not out of the question. However, apart from some speculation in the concluding section, I will limit my attention in Kant to the sort of "two-aspect" view that I have so far sketched with respect to the *empirical* distinction.

2 Transcendental phenomenalism (Kant)

Although I will be limiting my attention to the empirical distinction between things-in-themselves and appearances in Kant, the position that I have in mind may nonetheless reasonably be called a kind of "transcendental phenomenalism", and at least to that extent a "transcendental idealism". One reason for this is that it seems to me the most natural reading of what Kant is attempting to accomplish by way of his notion of the "transcendental object = X" in the first edition of the Transcendental Deduction in the *Critique of Pure Reason*.[8] With regard to the question of how "to make understood what is meant by the expression 'an object of representations,'" Kant says the following:

> [S]ince we have to do only with the manifold of our representations, and that *X* which corresponds to them (the object), because it should be[9] something distinct from all of our representations, is nothing for us, the unity that the object makes necessary can be nothing other

[8] This is not the only ground supporting the label "transcendental phenomenalism". Of particular importance, given Kant's proclaimed use of the term 'transcendental' more generally (A56–7/B80–1), is relevance to the possibility of *a priori* knowledge. This will come out at the end of this section.

[9] I assume Kant means: "insofar as it is supposed to be".

> than the formal unity of the consciousness in the synthesis of the manifold of the representations. (A105)
>
> All representations, as representations, have their object, and can themselves be objects of other representations in turn. Appearances are the only objects that can be given to us immediately, and that in them which is immediately related to the object is called intuition. However, these appearances are not things in themselves, but themselves only representations, which in turn have their object, which therefore cannot be further (*nicht mehr…kann*) intuited by us, and that may therefore be called the non-empirical, i.e., transcendental object = *X*. The pure concept of this transcendental object (which in all of our cognitions is really always one and the same = *X*) is that which in all of our empirical concepts in general can provide relation to an object, i.e., objective reality. (A108–9)

One might be inclined to assume that the concept of "formal unity" in the first of these passages, and that of the "transcendental object = *X*" in the second, are simply Kant's way of referring to whatever it takes to regard a manifold of appearances as all appearances of one object, given that "we have to do only with the manifold of our representations". But both here directly and in light of other things that Kant says, it seems to me much more natural to suppose that he is concerned with whatever in general it takes to regard a manifold of appearances – each of which might indeed be regarded as an appearance of an object – as from another point of view *itself one* object, thereby in a manifold of appearances. Here we may note, first, that he begins the second passage precisely by noting that appearances can be appearances *of* an object ("have their object") and yet also be regarded as objects in their own right. Second, he notes in the same passage that the object *of* the appearance in question cannot be *further* "intuited", beyond whatever intuition is to be found precisely *in* the appearance of it in the first place; thus the suggestion would seem to be that, just by virtue of the latter, it must indeed likewise be legitimate to think of the *object* as thereby "intuited". Third, this seems to me the most natural reading of Kant's reference to a purely "formal unity of consciousness in the synthesis of the manifold" in the first passage. Whatever sort of unity of consciousness is involved in regarding a manifold of appearances, despite their distinctness, all as appearances of one object, that would hardly be a merely "formal" matter. But the sort of unity of consciousness expressed in a reference to: *X*, *X*, *X*, and so on, throughout some manifold of appearances, might well be so described.

Confirmation of this reading can be found in the following:

> **Sense** represents the appearances empirically in **perception**, the **imagination** in association (and reproduction [i.e., in the eventual arrival of appearances associated with given ones, insofar as *recognition of an object* might thereby be effected]), and **apperception** in the **empirical consciousness** of the *identity* of these reproductive representations with the appearances through which they were given, hence in **recognition**. (A115; italics and bracketed gloss are my own; bold text in original.)

Kant does not simply speak of the "identity" of an object *through* a manifold of appearances "reproductively" associated with a given appearance, nor does he speak of mere "synthesis" or "connection" of the former with the latter. He speaks of an identity of the antecedent with the subsequent *appearances*. They are, presumably, from the point of view in question *one and the same object*, despite being from another point of view a manifold of appearances. Thus the notion of the "transcendental object" – in all of our cognitions "always one and the same = *X*" (A109)[10] – is arguably Kant's way of introducing what I have called "transcendental phenomenalism".

A number of other considerations point to the same conclusion. The following stems from Kant's argument, in the "Analogies of Experience", that we need to grant a priori status to a principle regarding the permanent existence (substantiality) of matter:

> Our *apprehension* of the manifold of appearance is always successive, and is therefore always changing. We can therefore never determine from this alone whether *this manifold* [my emphasis], *as object of experience* [my emphasis], is simultaneous or successive, if something does not ground it *which always exists*. (A182/B225)

This says that we need to appeal to the principle of permanence, because we need to be able to judge, with respect to what from one point of view is a multiplicity or manifold of appearances – for example (an example that Kant eventually offers in the Second Analogy), the successive

[10] Cf. A253: I "have no concept of it except merely that of the object of a sensible intuition in general, which is therefore the same for all appearances". If it is not clear from this passage, it is certainly so for A109, that what is being thought of as "one and the same" is the *object* in question, not the *concept* of it.

appearances of a house perceived from different perspectives – whether or not it is in fact, "as object [*Gegenstand*] of experience", a manifold at all (as opposed, for example, to a single thing successively apprehended). Thus what seems to be in question is the possibility of judging with respect to each member of a manifold of appearances that it is *both* from one perspective an appearance that is successive to or preceding others *and*, from another perspective, an object appearing *through* such appearances.

Now one might suppose that Kant is simply being careless. Setting aside the possibility of sheer hallucination, or "the mere effect of the imagination (in dreams as well as in delusions)" (B278),[11] we might suppose that the manifold of appearances of which there is successive apprehension is simply either one, or more than one, really existent thing *in* a certain succession of "appearings", or one or more than one really existent thing *as* it (or *as* they) are successively appearing. And when Kant speaks of determining whether this manifold "as object of experience" is simultaneous or successive, he is simply speaking of determining whether or not the thing or things appearing *in* this manifold is (or are) an instance of a succession of things. But it would surely have been an easy matter for him simply to have said that we "can therefore never determine from this alone whether *the object of* this manifold (or the object appearing therein) is simultaneous or successive".

In the following passage Kant also speaks exactly as we would expect him to speak if he were espousing, and as would seem at least odd were he not espousing, transcendental phenomenalism:

> Here that which lies in the successive apprehension is considered as representation, but the appearance that is given to me, in spite of the fact that it is nothing more than a sum of these representations, is considered as *their object* [emphasis added]. (A191/B236; cf. A370)

[11] Cf. *Reflexion* 5653, "Against material idealism": "inference from the immediate consciousness of mere representations of things outside us to their existence, which inference, however, is not self-evident in its conclusion, as is proven by the well-known property of our imagination, which is a faculty for intuitively representing objects even without their presence" (Kant 2005: 281; *Ak*. 18.306). (Throughout, references of the form "*Ak*. x.y" are to volume (x) and page (y) of the "Academy Edition" of Kant's works (Kant 1902–). Since the Academy pagination is included in Kant (2005), I refer subsequently to the *Reflexionen* only in that mode.)

Again, what from one perspective is a manifold of representations *of* some appearing object, from another perspective *is* that very object.

The transcendental phenomenalist reading also helps with the following passage, frequently considered objectionable in Kant. It occurs as part of the *Prolegomena's* distinction between "judgments of perception" and of "experience":

> All of our judgments are at first [*zuerst*] mere judgments of perception; they hold only for us, i.e., for our subject, and only afterwards [*hintennach*] do we give them a new relation, namely to an object ... [12]

This may seem to imply that all of our judgements about ordinary objects of perception are at first judgements with respect to how they appear in perception. That would conflict with Kant's insistence that we are immediately aware of bodies in space and don't need to achieve that awareness by way of the awareness of something else from which they would have to be inferred.[13] But for the transcendental phenomenalist, the statement is in order. On that view, all our judgements about ordinary appearing objects are, from another perspective, judgements about "items" that theoretically might not be, or might not have been, such objects at all. For (allowing again for the possibility of "the mere effect of the imagination"), those objects might be, or might have been, at most *possible* such objects, or possible appearances thereof. Furthermore, on the view in question, there is no reason why an appearance really being from the appropriate perspective either an ordinary appearing object, or the appearance of one, should be regarded as a necessary condition for one's having it as a possible *candidate* for such determinations in the first place. But, on that same view, having such an item as at least a possible candidate for such determinations is a necessary condition for judgement regarding ordinary appearing objects. Thus in at least that sense, on the transcendental phenomenalist view, there is a straightforward respect – yet still compatible with a "direct realist" theory of perception – in which judgements about ordinary appearing objects must "at first" be judgements with respect to something "subjective".

---

[12] (Kant 2002[1783]: § 18, p. 92 [*Ak*. 4.298]). (Since the Academy pagination is included in Kant (2002), I refer subsequently to the *Prolegomena* only in that mode.)

[13] Cf. the first (A366–7ff) and second edition (B274ff) "Refutation(s) of Idealism".

Transcendental phenomenalism also provides a natural account both of Kant's original and seemingly subjectivistic "refutation of idealism" and of his decision to rewrite it in the second edition of the first *Critique*.[14] The crux of the original refutation was that, since bodies are just "a species of [our] representations", we are entitled to insist, in the face of sceptical urgings to the contrary, that we are immediately *aware* of bodies, and not of intervening items from which we might infer them:

> [M]atter [for the transcendental idealist] is only a species of representations (intuition)... external objects (bodies) are merely appearances, hence also nothing other than a species of my representations, whose *objects* are something only through these representations, but are nothing separated from them. (A370; my emphasis)

This might seem to give voice to the subjectivist view that bodies are mental entities or reducible to mental entities. But on the transcendental phenomenalist view, it would very naturally just mean this: to grant immediate awareness of what are at least the possible appearances of ordinary appearing objects is by that very fact to grant those "items" the status of candidates for regard precisely *as* ordinary appearing objects. Similarly, immediately after announcing the intention to subject his opponents' premises to critical examination, Kant concedes that "only what is in ourselves can be immediately perceived" (A367).[15] That may sound objectionably subjectivistic. But on the transcendental phenomenalist view it would very naturally just be taken to say that any object of which we are capable of immediate perception must be an object of which it is always correct, from at least the "subjective" perspective, to regard as the mere appearance *of* an ordinary appearing object.

Assuming that this was in fact Kant's thinking in the original argument, we might then explain the inadequacy that he came to perceive in it, without supposing it to have involved undue subjectivism. The

---

[14] See previous note. Kant himself says there is no difference except in "mode of presentation" (Bxxxviii).

[15] To be sure, Kant goes on to concede two other things that might suggest either that he has not yet in fact begun his critical examination or else has more to recant in the second edition, namely, (a) "that my own existence alone could be the object of a mere perception" and (b) that "the existence of a real object outside me... is never given directly in perception". But with the qualifications "a mere (*einer blossen*)" in (a) and, in (b) but above elided, "if this last word [i.e. 'object'] is taken in an intellectual signification", such suggestions are unnecessary.

point would be one that Kant in fact emphasized in the second edition version of the argument, namely, that it is always in principle possible that an appearance that is a candidate for recognition as an ordinary appearing object – and not merely as an (at least possible) appearance of one – might possibly be a mere hallucination or delusion (B278).

He makes a similar point in the *Prolegomena*:

> If an appearance is given to us, we are still completely free as to how we want to judge things from it. The former, namely the appearance, was based on the senses, but the judgment on the understanding, and the only question is whether there is truth in the determination of the object or not. The difference between truth and dream, however, is not decided through the quality of the representations that are referred to objects, for they are the same in both ... (Remark III to Part One, Kant 2002[1783]: 85; 4.290)

Thus no simple inference is possible from the fact (emphasized in the first edition) that perceivable bodies are the very same "items" that we are capable of perceiving as (at least possible) appearances *of* bodies, to the conclusion that we are immediately aware of actually existing bodies. The second edition speaks to this by both emphasizing the possible objection and providing independent reason for holding that in any case not *all* appearances that are candidates for determination as actual bodies could fail to be such bodies.

It is also useful to note that a different formulation of the Refutation, formulated on several occasions in unpublished *Reflexionen*, seems to make sense only in terms of the doctrine of transcendental phenomenalism. There, as in the *Critique*, Kant rejects the supposition that we possess only an "inner", and not also an "outer" sense, that is, a sense whose immediate objects would be bodies in space. But what he adds to the way he puts it in the published version of the argument is that the rejected alternative would be tantamount to transforming the very space of bodies *into* the "form of inner sense". That is, on the rejected alternative, "space itself would be time" (*Refl*. 5653; 18: 310); "the representation of space would be transformed (*verwandelt*) into a representation of time, i.e., it would be possible to represent space as a time" (*Refl*. 6311; 18: 611).[16] The supposition thus seems to be that we are given bodies in space only to the extent that what we are given – those very

[16] Cf. *Refl*. 5655 (18.314–15) and 6315 (18.618–19).

appearances – would all be, under a certain counter-factual supposition, no more than "inner" appearances.

Transcendental phenomenalism is also arguably implicit in Kant's distinction between the "mathematical" and "dynamical" categories, and the corresponding a priori principles. Kant says that the mathematical categories and principles are "concerned with objects of intuition (pure as well as empirical)", while the dynamical are "directed at the existence *of these objects*" (B110; my emphasis); the former "pertain merely to the **intuition**", the latter "to the **existence**", of appearances (A160/B199; Kant's emphases). Thus specifically, the Analogies of Experience "do not concern the appearances and the synthesis of their empirical intuition, but merely their **existence** and their **relation** to one another with regard to this their existence" (A178/B220; Kant's emphases).[17] Otherwise put, while the mathematical principles pertain to appearances "with regard to their mere possibility", that is, with respect to "their intuition and the real (*dem Realen*)[18] in their perception" (A178/B221), the Analogies "bring the existence (*Dasein*) of appearances under rules *a priori*" (A179/B221). Thus:

> That something happens, therefore, is a perception that belongs to a possible experience, which becomes actual if I regard the position of the appearance as determined in time, thus if I regard *it* as an object that can always be found in the connection of perceptions in accordance with a rule. (A200/B245; my emphasis)

One may of course object to reading any of this in terms of transcendental phenomenalism. For there may seem to be no need to suppose that the appearances (or intuitions) in question, in addition to being candidates for determination as ordinary appearing objects, or even as the appearances *of* such objects (at least in the sense of being such objects qua appearing), are also at least in principle candidates for determination as *merely possible* objects (or appearances thereof). Why not simply suppose that the appearances in question – those that are candidates for determination as objects with respect to their actual existence – simply *are* just ordinary appearing objects? In that case, Kant would only be

---

[17] Cf. *Prolegomena*, § 25 (4.307), § 26 (4.309–10).

[18] Here, of course, "the real in their perception" (*dem Realen ihrer Wahrnehmung*) does not mean actual as opposed to possible existence (*Dasein*), but simply refers to that in an appearance which, corresponding to sensation, always has a certain "intensive magnitude" or degree of intensity (A166ff/B207ff; cf. A143/B182).

saying that the mathematical principles are grounded in a concern with the possibility of objects qua apparent in intuition, but without regard to conditions for judgement as to their actual existence; the dynamical are concerned with just those conditions. But the difficulty with that is that Kant would then either not be saying anything positive – as he seems to want to[19] – as to the actual import of the mathematical principles, or else his formulation is simply confused.

This leads finally to what seems to me the main reason for insisting that Kant in any case *ought* to have endorsed transcendental phenomenalism. The reason is simply that this would allow him to hold both that the objects with whose existence the principles of understanding are concerned are accessible without the need for *intervening* objects, and also that the determinations expressed by those principles do apply to ordinary appearing objects, and not simply to the ways in which they appear.

It might to be sure seem obvious that Kant's insistence that the principles in question do apply to ordinary appearing objects is just his way of insisting that they bear on how those objects are *constrained* to appear (and how we are constrained in our thinking to represent their appearing), as opposed to how they merely happen to do so. That, one might argue, is just what it means to say that those principles apply to objects considered (empirically) "in themselves". In support of this, one might, for example, note Kant's distinction between propositions regarding tastes and colours, on the one hand, and regarding the space within whose "intuition" the mathematical principles are supposed to be grounded, on the other:

> Space, on the contrary, as a condition of outer objects, necessarily belongs to their appearance or intuition. Taste and colors are by no means necessary conditions ... [they are] contingently added ... Hence they are not *a priori* representations. (A28–9)

---

[19] Thus the mathematical principles are said to teach "how both their [appearances=] intuition and the real in their perception could be generated in accordance with rules" (A178/B221); cf. *Prolegomena*, § 26 (4.309–10): the mathematical principles "refer to the genesis (*Erzeugung*) of intuitions". In any case, Kant's emphasis on independence from questions of existence is presumably meant to refer back to his emphasis on the fact that the mathematical principles are grounded in our positive ability to learn precisely *from* intuiting something "prior to the existence" of the objects of the intuition in question (A26/B42).

In reply, however, I would note first that in revising this passage for the second edition Kant dropped his formulation in terms of the distinction between what is necessarily or only contingently a condition for appearances. What he left is just the point that our representation of space, unlike that of tastes and colours, is a source of synthetic a priori propositions (A30/B45). But that is offered, as it was in the original version, in order to explain why a special sort of *ideality* is to be attributed to space, not a special sort of (namely, "empirical") reality.[20] Second, in response to the objection of "insightful men" regarding his view of time, complaining that it degrades temporal reality to mere appearance, Kant does not counter in terms of what it means to say that something is or is not (empirically) "real". Nor does he introduce the distinction between necessity and contingency with respect to appearances. He simply points to the fact that we do in fact *have* a "representation of time", and that the contents of our "inner awareness" do in fact appear *in* what is thereby represented (A37/B53–4). In other words – as in the first edition Refutation of Idealism – the primary point is just that, with space and time as "representations" encompassing spatio-temporal appearances, Kant already regards himself as free from the charge of denying the possibility of immediate access to *the very objects* that his opponents have been taking for empirical realities. In any case, to repeat: transcendental phenomenalism at least makes it possible to maintain that, if the a priori principles in question apply to "appearances", then they likewise apply to ordinary appearing objects, and not merely to the way in which (necessarily or contingently) they appear. What's more, it makes it possible to do

---

[20] In both editions Kant also says that "things like colors, taste, etc., are correctly considered not as qualities of things but as mere alterations of our subject, which can *even* (*sogar* [my emphasis]) be different in different people". The use of the term *even* suggests that the fundamental point was not captured with the original observation regarding the relative necessity or contingency of various ways in which objects appear to us. And in concluding the discussion in both editions Kant in fact says that:

> The transcendental concept of appearances in space, on the contrary, is a critical reminder that absolutely nothing that is intuited in space is [from the transcendental perspective] a thing in itself, and that space is not a form that is proper to anything in itself, but rather that objects in themselves are not known to us at all. (A30/B45)

The "transcendental concept" of appearance is thus not offered as a "critical reminder" that objects of human experience are necessarily spatial, but not necessarily coloured, odiferous, etc.

this without commitment to a subjectively reductionist view of the objects in question. And this is something that above all Kant wanted. Given the essential connection between "transcendental" considerations and the possibility of a priori knowledge (A56–7/B80–1), this reinforces adoption of that term for the particular variety of phenomenalism that I have been ascribing to Kant. (At the end, in connection with the broader notion of transcendental "idealism", I emphasize a further consideration in support of that terminology.)

It of course remains to consider how to make sense of this approach. In particular, we may seem to be confronted with the following difficulty. (a) There needs to *be* something antecedently available for alternative determinations as either empirically real phenomenal existents or at least possible appearances thereof. But (b) if there *is* something antecedently available for such determinations, it would seem to have to be something "in itself" on a level transcending that on which it might eventually be regarded as a "thing in itself" in the empirical sense; this may seem to threaten the ontological "multivalence" required by transcendental phenomenalism of the basic items (appearances) to which empirically real phenomenal existents are supposed to be reducible. I will return to this issue in Section 4.

3 Transcendental phenomenalism (Sartre)

I have so far not provided much of an argument to show that Sartre endorses any sort of phenomenalism at all. I have noted his claim that, whatever their limitation in advancing the progress of "modern thought", those whom Sartre himself calls phenomenalists have "justifiably (*à juste titre*) reduced the object to the connected series of its appearances" (21/26). As we have seen, he also refers to the modern "theory of the phenomenon" as "ours", and says that "We shall interpret this by saying that the series of [the object's] appearances is connected by a principle which does not depend on my whim" (5/13). On the other hand, in the same passage he also speaks of taking "our" theory as having "replaced the *reality* of the thing by the *objectivity* of the phenomenon", and I have suggested that it is unclear how we might in fact be able to take this, consistently with supposing that the "replacement" in question is something with which Sartre himself really means to agree. Certainly, it will not be acceptable to Sartre to ground any such undertaking with nothing more than an "appeal to infinity" (*ibid.*), that is, an appeal to the infinite multitude of the appearances to which any phenomenon would be

"reduced".[21] Minimally, by viewing any appearance as an "impression" in the sense of a "subjective plenitude" (22/27), such an approach fails to accord sufficient ontological status to appearances from the start:

> Being has not been given its due. ... [22/27] to be conscious *of* something is to be confronted with a concrete and full presence which *is not* consciousness. ... If then we wish at any price to make the being of the phenomenon depend on consciousness, the object must be distinguished from consciousness not by its *presence* but by its *absence* ... This is the appeal to the infinite of which we spoke in the first section of this work. For Husserl, for example ... truly objectifying intentions are empty intentions, those which aim beyond the present subjective appearances at the infinite totality of the series of appearances. ... But how can non-being be the foundation of being? ... [23/27] It is true that things give themselves in profile: that is, simply by appearances. And it is true that each appearance refers to other appearances. But each of them is already in itself alone [*à elle toute seule*] a *transcendent being*. (21–3/26–7)

Now I do not intend to argue – or rather: I will in fact argue, but with an important qualification – that the being of appearances gets sufficient due for Sartre simply by way of the transcendental phenomenalist approach to appearances, not simply as candidates for inclusion in inexhaustible

---

[21] Cf. Gardner (2009: 53–4): "Sartre endorses Husserl's 'appeal to the infinite', in the sense that he agrees that it captures what it is like for consciousness to have an object which possesses objective being ... [but] presses Husserl on is the question of what makes this structure *possible* ...", namely, "the *trans*-phenomenal being of the object". This in turn, as earlier explained, requires that "every appearance would involve the *phenomenon of its being*" in the sense that, for example, "not only does *the table* appear to me; in addition, there appears to me *the being of* the table" (41). I do not find it clear whether or not this should be taken as an endorsement of, or at least as meant to be compatible with, transcendental phenomenalism. If only for that reason, as for the ontological status of appearances *as such* – that is, of appearances insofar as they are not regarded as real existents – I have not been able to get a clear sense as to Gardner's position. But if the passage just quoted in fact contains an endorsement of transcendental phenomenalism, then I take the point to be that sufficient ontological due is given to appearances as such just by the fact that they are at least *possibly* real existents. (As I argue in Section 4, our position on this question has to turn on a reading of Sartre's view, in the chapter on "Transcendence", of the "ground" (*fond*, *fondement*) of "being-in-itself" from which appearances as such "emerge", about which Gardner (108–9) does not seem to me to say enough to indicate a position.)

series of appearances/profiles of phenomena, but rather as some sort of "items" that are by that very fact at least possible phenomena in their own right, that is, at least possible empirically *real existents*. There is (again pending the qualification to be introduced) no need to regard appearances *as such* (as opposed to phenomenal existents as such) as having, in any further sense, a being "in-themselves". Given such a view of appearances as such, one might perhaps then also suppose that the reality of phenomena might after all be replaced by the latter's "objectivity", insofar as that objectivity is explicated in terms of infinite series of appearances so regarded.[22] However, whatever the eventual verdict regarding appearances as such, I take Sartre's position at least to include the view that a real existent has an irreducible being-in-itself, although it of course remains to be seen how this might be squared with any sort of "phenomenalism". In any case, given that Sartre claims to see progress in the modern idea of the "reduction" of phenomena to appearances, and given, as I will argue, that transcendental phenomenalism is compatible with the irreducible "being-in-itself" of phenomena, I will regard it as reasonable simply to proceed along the suggested line, and then turn to the question both of its ultimate intelligibility and of its compatibility with Sartre's view of being-in-itself.

The first page of "The Pursuit of Being"[23] provides two reasons for supposing that, whatever "modern thought" might mean by reducing a phenomenal existent to the series of appearances which manifest it, this is not only not meant to exclude, but is meant to entail, regarding each of the relevant appearances as also itself a phenomenal existent. First, Sartre offers there, as examples of "*apparitions qui manifestent l'existant*", (a) a series of "accelerations, deviations, etc.", which manifest some particular physical force and (b) various "physical-chemical actions" such as "electrolysis, the incandescence of a carbon filament, the displacement of the needle of a galvanometer, etc.", which manifest an electric current. Thus that to which the *reducenda* are supposed to get reduced are themselves series of items describable as ordinary phenomenal existents.

In the same paragraph, Sartre also says something else to what I take to be the same effect. Speaking of the various appearances that manifest an electric current, and having pointed out that none alone is

[22] When Sartre says "*Concevons bien, en effet, que notre théorie du phénomène a remplacé la réalité de la chose par l'objectivité du phénomène*" (5/13), he might be taken as saying only that this is the most on offer so far, short of his own view, remaining to be articulated. And he does immediately "interpret" the claim in question in a fairly neutral way, in terms of series of appearances "connected by a principle which does not depend on my whim".

[23] That is, p. 1 of the French edition; pp. 1–2 of Hazel Barnes's translation.

sufficient to reveal it, nor does any indicate anything "behind itself", he says that each of them "indicates only itself and the total series" (*indique elle-même et la série totale* (4/11)). Why would Sartre speak of an appearance as "indicating" *itself* in this context? (This same locution is also present at several other points in "The Pursuit of Being".[24]) At least along the lines of the "modern thought" in question, why not simply observe that, in apprehending an appearance as that *of* some phenomenal existent, rather than taking it to "indicate" the latter as a thing-in-itself behind itself, one would only be taking it to indicate a potentially infinite series of appearances to come (and, one might add, already to *have* come)?[25] In what sense would there be any question of

---

[24] "[T]he phenomenon exists only qua appearance; that is, it indicates itself on the foundation of being" (9/16). I return later to the language of being as a "foundation" for appearances as such, as opposed to something that – apart from their being taken as phenomena – they might be said to "have". In any case, some additional passages:

> Thus [by way of the reductionism of "modern thought"] we arrive at the idea of the phenomenon such as we can find, for example, in the "phenomenology" of Husserl or of Heidegger – the phenomenon or the relative-absolute. ... "to appear" supposes in essence somebody to whom to appear. But it ... does not point over its shoulder to a true being which would be, for it, absolute. What it is, it is absolutely, for it reveals itself as it is. ... it is *absolutely indicative of itself*. (4/12; Sartre's emphasis)
>
> [T]he first consequence of the "theory of the phenomenon" is that the appearance does not refer to being as Kant's phenomenon refers to the noumenon. Since there is nothing behind the appearance, and since it indicates only itself (and the total series of appearances), it can not be *supported* by any being other than its own. (6–7/14; Sartre's emphasis)

[25] Cf. Holmes (1984: 401–2):

> Initially objects ... are given as such immediately by virtue of their always pointing beyond themselves toward further "seeing" of themselves and the system of objects within which they are found. For example, the glass of wine appears with the character of an object which exists independently of my consciousness of it and it stands out as such from a background of ... all objects of whatever type or description.

And, more generally, the following (speaking of Husserl but to bring out a similarity with Sartre) might seem to suggest a transcendental phenomenalist approach:

> Objects present themselves not only as surpassing the consciousness of them but as belonging to the system of all objects of whatever type or description. Whether presented as real, imagined, remembered, anticipated, ideal, or however, each is an object of possible multiple consciousnesses of it. Each belongs to the world, in the broadest possible sense ... "World" here signifies the system of all actual and possible objects which correlates with the system of all actual and possible consciousnesses. (401)

However, see footnote 27.

self-indication on the part of the given appearance? But of course for the transcendental phenomenalist, insofar as what is in question is the appearance *of* some particular existent, that very same appearance is *ipso facto* also rightly regardable *as* the existent in question.[26]

In the following passage, Sartre is formulating a version of the modern view containing an element that I take it, again, he does not accept, namely, that of having "replaced the *reality* of the thing by the *objectivity* of the phenomenon ... [basing] this on an appeal to infinity" (5/13) – such "replacement" in turn supposedly required by the fact that each appearance is just on its own "an intuitive and subjective plenitude" (5–6/13). But it is striking how Sartre formulates the view:

> [T]he result is that on principle an object posits (*pose*) the series of its appearances as infinite. Thus the appearance, which is *finite*, indicates itself (*s'indique elle-même*) in its finitude, but at the same time in order to be grasped as an appearance-of-that-which-appears, it requires that it be surpassed toward infinity. (6/13)

Striking is not simply the notion of "self-indication" again. What is striking is also that the self-indicating item in question, which "posits" the series of its own appearances, is described as also itself a *given* appearance: as "an appearance-of-that-which-appears (*apparition-de-ce-qui-apparaît*)". And just a few lines further on the same page:

> What appears in fact is only an *aspect* (*aspect*) of the object, and the object is altogether *in* that aspect and altogether outside of it. It is altogether *within*, in that it manifests itself *in* that aspect; it [i.e., the *object*] indicates itself as the structure of the appearance, which is at the same time the principle of the series.[27]

---

[26] On the other hand, I take McCulloch to go too far in the generalization that "[w]hen [Sartre] talks of 'appearance', then, he is talking about the perceptible material world", as likewise in his conclusion from this: "That means, in this context, that he is arguing against Husserl's (and anyone else's) transcendental idealism" (McCulloch 1994: 103). It is quite a different matter, I take it, when Sartre talks of 'phenomena'.

[27] *il s'indique lui-même comme la structure de l'apparition*; Barnes translates, "shows itself." Also to note: the object/phenomenon indicates itself as the structure *of a given appearance*.

Cf. from the chapter on "Transcendence":

> [T]he yellow of the lemon is not a subjective mode of apprehending the lemon, it *is* the lemon. And it is not true either that the object X appears

As we might also note in connection with a passage discussed earlier, where Sartre objected to an attempt, by way of the absence of infinitely many appearances of the object, to combine consciousness-dependence of the being of the object with the object's distinction from consciousness, he there described the absence in question precisely as the *object's* absence ("*il faut que l'objet se distingue de la conscience non par sa présence, mais par son absence*" (22/27)).

Here is another passage. In it, Sartre concludes his second[28] response to the question as to how far, if at all, he might be able to join modern phenomenalists in reducing an object to the multiplicity of its appearances, or to the multiple ways in which it might be perceived. We have already seen part of his criticism. Here is something else that he says in the matter. The philosophers in question have mistakenly supposed that they have reduced the being of an object to some sort of *succession*. But it turns out that they could at most – and therefore circularly – have "reduced" it to the succession of *its own modes of being*:

> [T]hey believed they had reduced its being to the succession of its modes of being (*la succession de* ***ses*** *manières d'être*). That is why they have explained it by concepts which can be applied only to modes

---

> as the empty form which holds together disparate qualities. In fact the lemon is extended throughout its qualities, and each of its qualities is extended throughout each of the others. It is the sourness of the lemon which is yellow, it is the yellow of the lemon which is sour. We eat the color of a cake, and the taste of this cake is the instrument which reveals its shape and its color to what we may call the alimentary intuition. (257/222–3)

Although she introduces it in the context of the question of "being-in-itself", and does not relate it to the issue of phenomenalism, Hazel Barnes emphasizes this point: "I confront [the being of the lemon] directly. All of the lemon is present to me in each of its appearances to consciousness, even though it is not exhausted by its appearances" (Barnes 1992: 24). By contrast – and also by contrast with what he might be taken to say elsewhere (see earlier note) – Holmes seems to suggest a type of phenomenalism according to which an object would be a *mere* "totality of appearances":

> [W]hile holding a glass of wine I am paying attention to the balance of the glass and then wonder about the wine's taste. ... it is essential to what it means to be a series of appearances that they are all absent except for the one to which attention is now directed. ... Think of the glass of wine. The present appearing is seen as but one of a series, the rest of which are *not* now appearing. (Holmes 1984: 399)

[28] Earlier: 5–7/13–14.

> of being (*à des manières d'être*), for they are pointing out the relations between a plurality of already existing beings. (21/26; my emphasis)

Now one might suppose that the "plurality of already existing beings" is supposed to be a plurality of appearances of some object, each of them now regarded by Sartre as something "in itself", independently of eventual identification as (or as an appearance of) a real existent. But apart from the fact that, perhaps *always* at least elsewhere, Sartre calls only *real existents* "existing beings",[29] he seems pretty plainly to be saying here that each item in the plurality in question is not simply an appearance of, but is precisely a *mode of being* of the really existent object itself. The most reasonable interpretation therefore seems to me to be that he is saying just this: each of those items is precisely a really existent object, in some particular manner of its appearing. Consider after all what he says at the end of the next section, regarding the "transphenomenal being of phenomena":

> [T]he transphenomenal being of phenomena [is] not a noumenal being which is hidden behind them. It is the being of *this table*, of this *package of tobacco*, of the *lamp*, more generally the being *of the world* which is implied by consciousness. (24/29; my emphases)

Sartre's central concern, it seems, is precisely with giving sufficient due (compatibly with some sort of modern "phenomenalism") to the being-in-itself of *phenomena*, or really existent things such as tables and packages of tobacco. So far, at least, there is therefore no reason to suppose that what is in question is the need to regard appearances *as such* as anything "in themselves". Or at least as I would conclude, there is so far no reason to suppose a "transphenomenal being" for appearances as such *beyond* the need to regard them as at least possibly real existents.

But then what about the passage cited earlier in which, speaking of a table, Sartre insists that "there is a being of the thing perceived – *as perceived* (*un être de la chose perçue en tant qu'elle est perçue*" (18/24)? This in fact raises precisely the point of the preceding example. If the claim is not that appearances as such have a being of their own, then what is the point of emphasizing that there is a being of the table *as perceived*? If we are meant to suppose that at least the table has a being of its own, and setting aside the *trivial* sense in which the claim in question would

[29] Possible exception: speaking of *being itself* as "existing" ("that the being of that which *appears* does not exist *only* in so far as it appears" (24/29))?

therefore be true, it may seem that there is no other alternative. But there is. The alternative is transcendental phenomenalism. Again:

> *Transcendental Phenomenalism*. Any judgement of an appearance *A*, to the effect that it is a phenomenon (a real existent), is a judgement to the effect that *A* is, from one point of view, a phenomenon and also, from another point of view, an appearance within an infinite series of appearances *of* the very phenomenon in question.

It seems to me that this is reasonably regarded as entailing that there is a being of the thing perceived *as perceived*, but it is neither a trivial claim nor does it grant being-in-itself to appearances as such (except in the sense that they are at least possibly *phenomena*). The same point might therefore also be regarded as applying to the following:

> It is true that things give themselves in profiles: that is, simply by appearances. And it is true that each appearance refers to other appearances. But each of them is already in itself alone a *transcendent being*, not a subjective material of impressions. (23/27)

This reading might also be regarded as confirmed by the following, as Sartre continues the passage previously quoted:

> Let us note first that there is a being of the thing perceived – *as perceived*. Even if I wished to reduce this table to a synthesis of subjective impressions, I must at least remark that it reveals itself *qua table* through this synthesis, that it is the transcendent limit of the synthesis, the reason for it and its end.

Of course Sartre denies that we could in fact coherently develop modern phenomenalism while regarding appearances as subjective impressions. For a transcendental phenomenalist, at least, so doing would amount to regarding subjective impressions as at least possible tables and packages of tobacco, and not merely as at least possible appearances *of* tables and packages of tobacco. The absurdity of this is supposedly what drives Husserl to seek escape by way of an appeal to infinity. In any case, and again under threat of triviality, what is the point of insisting – *given* that what is in question is "reducing the existent to the series of appearances *which manifest it*" (3/11; my emphasis) – that the table, through (*à travers*) the synthesis of appearances in question, reveals itself precisely *qua table* (*en tant que table*; Sartre's emphasis)? The

most reasonable answer is presumably that each synthesized "item" is as legitimately qualified for regard as a table as for regard as an appearance *of* one.

To be sure, Sartre also speaks of the table as the "transcendent limit" of the relevant synthesis, as its "reason" and "end". This might suggest a more traditional form of phenomenalism than what I have suggested he is pursuing. But he at least does not suggest that such characterizations should be regarded as an *explication* of the claim that the table reveals itself qua table through the synthesis in question. And indeed, if they were so regarded, we would be again faced, at least in context, with the charge of triviality. A perfectly reasonable explanation for the terminology is simply that Sartre is trying to play a reasonably conscientious devil's advocate, given that he has chosen, for his dialectical purposes, to allow the infinitizing rescue effort voice just one more time (22–3/27–8; cf. 5–7/13–14).

As indicated earlier, however, I still need to qualify the suggestion that the only sense in which Sartrean appearances as such "have", or "lay claim to", a being of "their own" (7/14, 17/23) is that they are always at least possibly real existents. The point needs to be formulated more precisely in the terms that Sartre himself in fact prefers, namely – and setting aside the fact that, in the following passages, he is not always speaking of appearances as such but of *phenomena* – formulated not in terms of the idea of appearances *as* something in themselves, or as "having" a "being-in-itself",[30] but rather in terms of the idea that they are something for consciousness only against a *ground* or *foundation* of (or against a ground or foundation that *itself* "has" – this point to be pursued in the next section) being-in-itself.

Thus, speaking of that particular appearance which is "the phenomenon of being" – that is, which is being in its "appearance" as the very *meaning* (*sens*) of being[31] – Sartre describes it as "an appearance which, as such, needs in turn a being on the *foundation* [*fondement*] of which it can reveal itself" (8/15; cf. 9/16). But the point is derivative from a point about phenomena generally: "it requires, as phenomenon (*en tant que*

[30] To be clear, I do not want to deny the appropriateness of such formulations in the case of phenomena, i.e. real existents. But it would be inappropriate to suppose that *appearances* as such "have" being-in-itself in anything other than the sense that (a) they are at least *possibly* real existents (which must be regarded as being something in the sense of being-in-itself) and (b) they are always apprehended on the *ground* of something necessarily regarded as being in that same sense.

[31] See 25/29, quoted below.

phénomène), a foundation which is transphenomenal" (9/16); speaking explicitly of "being-in-itself" in the section so titled: "existents appear before consciousness on the foundation of their being ... being is the ever present foundation of the existent" (24/29); and speaking of any and every "this" (*ceci*): "it is what is revealed on the undifferentiated ground (*fond*) of being" (252/218).

Again, Sartre is speaking in these passages of "appearances" that he is also prepared to call "phenomena", not of what I have been calling appearances "as such", that is, appearances insofar as they are in principle determinable *as* phenomena. This might lead one to suppose it to be his view that, strictly speaking, appearances as such "have" being-in-itself, or *are* in themselves, in some more fundamental sense and that phenomena – insofar as they are "reducible" to appearances – more exactly have such being only as their ground or foundation. Of course, this would not explain Sartre's reference to an *undifferentiated* ground of being, which is a crucial notion to which I will return. Nor would it explain his focus on "the transphenomenal being of phenomena" precisely as "the being of *this table*, of this *package of tobacco*, of the *lamp*, more generally the being *of the world* which is implied by consciousness" (24/29; my emphases).[32] As we have also seen, while there is independent reason for ascribing a transcendental phenomenalist view to Sartre, ascription of an antecedent "being of their own" to appearances, conditioning their determinability as (at least possibly) phenomena, i.e. as real existents, or alternatively as appearances *of* phenomena, would seem to stand in the way precisely of such ontological multivalence.

Compatibly with transcendental phenomenalism, if appearances as such have any sort of being, beyond their being at least possibly real

---

[32] Of course the "phenomenon of being", as a particular "appearance", calls for special treatment. Presumably, it makes no sense to speak of an "appearance" of being in principle determinable *as* the phenomenon in question. That would amount to a case in which *being itself* would be directly present to us, just as a package of tobacco, *of which* it were the being, might be. But "the phenomenon of being" is not being itself as a phenomenon in the way we might say the phenomenon (of a) table is the table itself as phenomenon. Rather, as we have seen, the phenomenon of being is the *meaning* that being has for us. And it is only in that sense that being can "appear" to us:

> Consciousness can always pass beyond the existent, not toward its being, but toward the *meaning of this being*. ... The meaning (*sens*) of the being of the existent in so far as it reveals itself to consciousness is the phenomenon of being. This meaning has itself a being, on the foundation of which [Barnes: "based on which"] it manifests itself. ... The phenomenon of being is not being ... But it indicates being and requires it. (25/29)

existents, then it must indeed be at most in a sense in which they are something for consciousness only against, or only "on", a *ground* whose very meaning includes *its* having being-in-itself. This is the view that I will develop in what follows as in fact Sartre's, though not Kant's. As noted earlier, however, it will not exclude regarding *phenomena*, despite their being in some sense "reducible" to series of appearances, as having being-in-itself in a stronger sense. In particular, at least in Sartre, it will not exclude regarding phenomena as having being-in-itself in just the same sense as that which applies to the *ground* of appearances as such. By contrast, as I will argue, while the notion of an "undifferentiated ground" is as much Kantian as Sartrean, it does not have the same ontological weight in Kant, namely, as a ground whose very meaning includes being-in-itself. In any case, even this much is at least *compatible* with regarding phenomena, despite their being reducible to series of appearances, as having some sort of being-in-itself.

Before proceeding to develop these final points, however, we need to see how to make sense of the very sort of ontological multivalence, with respect to appearances as such, that is part of transcendental phenomenalism as I have so far characterized it. But of course this needs to be done compatibly with regarding the appearances in question as at least in some sense "available" *for* the alternative determination thereby left open.[33]

4 Kantian form of sensibility, Sartrean transcendence

The most appealing path to a grounding of ontological multivalence with respect to the objects of consciousness might be thought to lie

[33] A complication that would need to be more fully addressed is raised by Sartre's claim, in the chapter on "Transendence", that the notion of an *apparition* (as also that of an *abolition*) should not be strictly part of "ontological" discourse at all (282ff/243ff). I assume that what is in question here is not "appearances" in the sense with which we are concerned in this chapter, and with which Sartre is concerned in "The Pursuit of Being" and elsewhere in the book (also sometimes called *apparences*). What is in question is presumably rather what we might call the "availability", or the "making" available, *of* appearances in that sense. I limit myself to suggesting here that at least part of the point might amount to the claim that, while the appearances in the sense of concern to transcendental phenomenalism are in principle multiply determinable as suggested in this chapter, they are always "originally" made available as at least already in *some* way determined in the first place. (This would require qualification of the notion of an appearance "as such" as I have employed it in this chapter.) In any case, Sartre's own reflections at this point seem to turn mainly on the consideration that such an original "occurrence" could not be regarded, from a strictly "ontological" point of view, as taking place *within the time* of the *apparences* thereby made available to a perceiver. Cf. footnote 44 for a comparable point regarding Kant.

in some notion of multiple "content". Of course this need not require literal containment of "content" (as perhaps with the Scholastic notion, as apparently adopted by Descartes, of the containment of "objective reality") within instances of consciousness; one might rather regard the relevant content simply as (at least) a part of the intrinsic *character* of instances of consciousness.[34] But it is difficult to see, on this sort of approach, how to accommodate multivalence with respect to any *given* object of consciousness. In a way that I hope to bring out, what is needed is rather the notion of an instance of consciousness whose essential "form", or whose very *being*, is precisely that of directedness *as such*; any even minimally determinate element of "content" will then be something further. Quite unlike Kant, however, what I take to be, at least for our present purposes, a counterpart notion in Sartre is connected by the latter with his own commitment to "transcendent being" on the side of the object of directedness:

> All consciousness is positional in that transcends itself to reach an object, and it exhausts itself in this same positing. (11/18)[35]
>
> This means that transcendence is the constitutive structure of consciousness; that is, that consciousness is born *supported* by a being which is not itself. This is what we call the ontological proof. No doubt someone [will object to the conclusion]... But this objection can not hold up against an analysis of what Husserl calls intentionality, though, to be sure, he misunderstood its essential character. To say that consciousness is consciousness of something means that for consciousness there is no being outside of that precise obligation to be a revealing intuition of... a transcendent being. (23/28)

---

[34] Cf. Descartes's Third Meditation and, e.g. First Replies (Descartes 1984: 74–6). For a recent study of Descartes placing particular weight on this notion, see Carriero (2009: esp. ch. 3). For some comments relating the two notions of "content," see McDowell (2009: 55–6).

[35] Barnes translates *pour atteindre* as "in order to reach". That aside, in what follows it will become clear how I differ from Phyllis Sutton Morris's (Morris 1976: chs 1–2) explanation of the claim that consciousness is "exhausted" in the "relation" of intending, namely, that it *is* precisely that relation, relating human bodies and objects. I agree that Sartrean intentionality is not a "character" or "feature" possessed by instances of consciousness as some sort of items in their own right. But unlike Morris, I take it that instances of consciousness *are* indeed – despite their "nothingness" – something in their own right. Still, they are not states or conditions of a subject of which intentional directedness is a character or feature: not even an *essential* one. Rather they are an irreducible sort of *event* of which the very *form* is intentional directedness (and of which Sartrean "facticity" – at most *including* the human body on some level of its being – is the "matter").

It remains to clarify this notion before showing how it facilitates the view sketched at the end of the preceding section. To set the stage I simply note for now that, on the approach that I have in mind, so long as alternatives for (more or less) determinate specification of objects are effectuated with respect to the directedness of a given instance of consciousness, it will be appropriate (at least from a transcendental phenomenalist perspective) to regard them as alternatives precisely with respect to a given object. For they will by that very fact be alternatives with respect to a given *instance of directedness*.

What I have in mind in Sartre is the basic structure of any instance of "transcendence of facticity", insofar as it is understood as an instance of the directedness of consciousness with a certain sort of "form-matter" structure: transcendence as the form of *directedness* as such, facticity as the medium *through which* that directedness is effectuated.[36] But whatever benefit this might yield in the present context for Sartre, the situation might seem quite the opposite with Kant. For there the "form" to which I would draw a parallel is the form pertaining to any instance of Kantian *Anschauung* as such.[37] It is of course not infrequently supposed that no consciousness for Kant – or at least no object-directedness, and so no consciousness beyond bare animal "sensation" – is secured just by way of *Anschauungsform* (or *Form der Anschauung*). This is frequently supposed to follow, for example, from the blindness of Kantian *Anschauung* without concepts (A51/B75). But whatever sort of "blindness" is meant in the context of pronouncements of that sort, it is clear that Kant recognizes levels of non-conceptual consciousness beyond

[36] I take it that any discussion of the transcendence/facticity structure in Sartre will need to distinguish between such a structure on what might, employing Husserlian terminology, be called the "noetic" and "noematic" sides. As I take it is in line with Sartre's understanding of a "phenomenological" approach to ontology, he appears for the most part to regard the noetic side – or directly that of instances of consciousness as such – as describable only precisely through its "noematic" *counterpart in the world as apprehended through* instances of consciousness: aspects of "transcendence" in the world reflecting the "transcendence" that is the very being/non-being of consciousness as such; aspects of "facticity" in the world reflecting the facticity *through which* the transcendence of consciousness "makes there be" a world for consciousness in the first place. In Aquila (1998), I have tried to show how this two-fold notion of transcendence/facticity facilitates interpretation of Sartre's view of consciousness of other subjects.

[37] It will become clear presently why (albeit, among other reasons) I avoid the term 'intuition' here.

bare sensation.[38] It should also be clear that he speaks of *Anschauung*, and so by implication "form" thereof, in more than one sense.

The core of Kant's isolation of a notion of *Anschauung*, and its distinction from concepts, is centred in the distinction between particulars, on the one hand, and our capacity for describing or characterizing particulars in terms that might also be employed more generally. It would therefore be no surprise if Kant mainly used the term *Anschauung* to refer precisely to particulars as opposed to generalities, or to would-be "abstract objects", e.g. to refer to appearances, or phenomenal objects as such, or the space and time in which (as "pure" *Anschauung*) such things appear. But he also frequently speaks of objects *of Anschauung*, and of *Anschauung of* objects.[39] It is to this that our own attention needs to be directed, that is, to *Anschauung* in the sense of what might be called (but not presuming the necessary *absence* of "concepts") "intuit-*ings*", as opposed to what English translations generally call "intuitions", in the sense of the corresponding Kantian "intuit-eds".[40]

---

[38] E.g. Kant (1992[1800]: Intro. VIII [9.64–5]). See also Naragon (1990).

[39] Kant frequently uses the verb (*anschauen*) to speak of "intuiting" all sorts of things: appearances (A93/B125), the manifold of appearances, space and time, particular spaces (A524/B552), determinations or relations of things, objects (*Gegenstände*, A27/B43, A93/B125, A293/B350, A490/B518; *Objekte* A38/B55), things in themselves (at least possibly), things intuited "under" the pure forms of intuition (A50–1/B75), things or objects intuited in space and time (A30/B45, B147, A373, A490/B518), such as drops of water (A263–4/B319). And he frequently speak of "objects of" our intuition, specifically as of sensible intuition (Bxxvi [*Objekt*], A27/B43, A35/B52, A51/B75, A90/B122, A772/B800) or simply as of intuition (B71, A79/B105, B110, B148, B150, A326/B382, A428–9/B456–7, A538/B566, A444/B472).

The extent of Kant's own recognition of this "act-object" ambiguity is arguably reflected in the fact that (albeit very briefly (A94, 97)) he substituted the title "*a priori* synopsis" for what I take to be the notion relevant to our present concern. In what I take to be mainly a terminological difference from my own reading of Kant, Waxman in fact simply identifies what I am calling "form of intuition", in the sense relevant to our present concern, with "*a priori* synopsis" (Waxman 2014: e.g. 79ff). Of course we might ask: Why did Kant not retain the terminological distinction in the B-edition? A reasonable suggestion might be that he thought he was clarifying the distinction in other terms in the much-discussed footnote to B160, by distinguishing between "form of intuition" and "formal intuition". However, it remains unclear just how he in fact intended either of those notions to be understood.

[40] Here, terminologically (but not in agreement with his view of the necessary *presence* of at least some minimal conceptual content), I follow the practice, e.g. of Wilfrid Sellars; cf. Sellars (1968: 8).

In these terms, then, what I have in mind is what I take Kant to articulate, albeit all too quickly, at the beginning of the Transcendental Aesthetic of the *Critique of Pure Reason*. The starting point[41] for the human cognitive process, he there tells us, is a particular sort of intuiting effectuated through the medium of what he calls "sensation", the latter characterized as a particular sort of effect upon perceivers. What concerns us is the "form" of such intuiting. Noteworthy about the latter are two things. First, its "object" (or at least, so far considered, its "proto-object") is a potentially all-encompassing, but so far *indeterminate*, field for the eventual discrimination of objects (or at least apparent objects) of possible cognition. Second, the fact that *there is* such a field, available to perceivers for eventual objective determinations, is an expression precisely of the very form of the intuiting in question, insofar as it is effectuated through whatever medium it requires. Now as suggested, Kant's own main interest is precisely in the field itself, and in the possibility of determinations within it. In his rush to get to this, he fails to make certain points sufficiently clear with respect to the intuiting in question, relative to which there is any such "object" in the first place. In particular, he moves too quickly to a concern with the form of the *field*, with the possibility of discriminating "matter" within it, corresponding to the sensation through which it is "intuited" in the first place, and of course with the need for concepts in the service of eventual objective determinations. In any case, the starting point is said to be an intuiting *of* (an at least so far considered) "undetermined object" (*der unbestimmte Gegenstand*). Kant calls this object (in the singular) "appearance" (A20/B34). Within it, appearances (in the plural) then need to be discriminated as a condition for the possibility of objective determinations.[42]

---

[41] This is of course not meant in the sense of chronological priority.

[42] That what is in question is aptly viewed as a "field" is evident from the fact that the relevant intuiting, as a state of the perceiver effectuated through the medium of sensation, is assumed from the start to be effectuated through a *manifold* of sensation. If there is a "transcendental" ground for that assumption, and given that the corresponding intuit-ed is "appearance" in the sense of that intuiting's "undetermined object", that ground presumably lies in the demand that all determination with respect to appearance must indeed relate to a presupposed *field* of appearance. And as Kant in fact notes more specifically with respect to determining objects in space, apart from the employment of *concepts* to that end, "everything in our cognition that belongs to *Anschauung*" consists in mere "relations" bearing on the possibility of things as relatively in motion and as affecting one another in space (B66–8). In the conference presentation to which I refer in footnote 49, I have argued that the most sense is to made of Kant's account of the role of concepts toward the determination of appearances, in the

Once our attention is directed to the form of a Kantian intuiting, we should not fail to be struck by certain similarities to, and differences from, an instance of Sartrean consciousness, or the Sartrean "for-itself". First, it seems pretty clearly to be Kant's view that an instance of intuiting through the medium of (mere) sensation[43] is an occurrence[44] that incorporates sensation within itself, as what Kant calls its "matter" – and includes *nothing else* as part of its matter – but that the *form* of such occurrence, consisting in nothing other than some level of apprehension[45] *through* the matter in question, is altogether irreducible in terms either of relations among whatever might be contained within that matter or between what might be thus contained and whatever eventually determined as occupying the field thereby apprehended. To this extent, what is in question shares the ontological status of an instance of Sartrean "transcendence", as a certain sort of irreducible and "absolute event" that is in some way *something* but yet, to the extent just indicated, *nothing more* than the facticity that is therein "engulfed" (*englouti*) and – as engulfed precisely by something that is nothing more than *itself* – "nihilated" (*néantisé*): "engulfed and nihilated in the absolute event

---

Transcendental Deduction of the Pure Concepts of the Understanding, when it is seen to turn on the need for the establishment of connections – namely, those constituting "transcendental unity of apperception" – precisely *within a presupposed* unity of consciousness of an indefinitely extendible field available *for* such determination.

[43] We may leave it an open question whether there can be Kantian intuitings whose "medium" is not sensation at all, or at least not wholly sensation. In Aquila (1989), and again in the conference presentation to which I refer in footnote 49, I have argued that Kant at least implicitly recognized the need for the incorporation within instances of intuiting of imaginative anticipations (and what one might call "posticipations"), on a purely "Humean" or "animal" level, whose correlate within the perceptual field is then a non-conceptually apprehended order of possibilities, in principle "available" for subjection to that *intellectual* structuring which is for Kant a pre-condition for properly *conceptual* apprehension within that field.

[44] I simply note, in passing, first, the likely discomfort occasioned by the Kantian view that the mental "acts" or "occurrences" whereby a perceptual field is generated and eventually "determined" for a perceiver cannot be regarded as occurring *within the time* of that field, or of whatever is eventually determined therein. For a discussion, see Waxman (2014: 121ff). Second, we might in any case note that Sartre shares this Kantian position; see footnote 33.

[45] It might be supposed that the first-edition introduction of "synthesis of apprehension in intuition" (A98–100) implies that something more is regarded even for the most meager "apprehension". However, a closer look should make it clear that what is in question there is the possibility for the making of discriminations *within* an otherwise indeterminately apprehended perceptual field.

which is the appearance of the foundation or upsurge of the for-itself" (130/118). (Sartre shortly thereafter equates such *néantisation* with the *transcendance* that is the very for-itself itself: "precisely the nihilation which is the origin of transcendence conceived as the original bond between the for-itself and the in-itself" (134/122)).

There are of course crucial differences between Kantian intuitings and instances of Sartrean transcendence. I emphasize three. First, the very form of directedness of any instance of Sartrean transcendence includes directedness toward a *goal*. Sartre introduces this point in the context of a sort of "metaphysical" myth (and so not a part of strictly ontological discourse) according to which the "birth" of the transcendence of facticity consists in an abortive effort on the part of being-in-itself to be its own foundation. The upshot is the for-itself as a kind of *non*-being "engulfing" facticity in an instance of directedness oriented toward the goal of becoming, by way of engagement within the field thereby apprehended, precisely something in-itself itself as well, thus in-itself-for-itself.[46] For Kant, by contrast, the bare "engulfing" of sensation within an instance of the directedness of consciousness consists in nothing more than an instance of *apprehension*.[47]

The second and third differences are more directly relevant to our present concern. At least on the most basic level, to begin, and to the extent that Kant at least explicitly recognizes,[48] it is only Kantian "sensation" that corresponds to the facticity engulfed and nihilated within instances of Sartrean transcendence; in any case, Kantian "facticity" is something "subjective". By contrast, of central importance for Sartre is that the ontological status of facticity is that of being-in-itself.

The final difference follows from this last point, as conjoined with a further similarity. The similarity is that, for both, the "field" toward (or within) which consciousness is directed, simply by virtue of the pure form of intuiting or transcendence, is always in some way characterized in its appearance to consciousness by the facticity *through which* it appears in the first place – at least insofar as it does so within whatever

---

[46] See e.g. 127ff/116ff, 133ff/121ff, 785ff/665ff.

[47] On might argue that, even in Kantian terms, it *should* involve more than this, namely, at least some sort of goal of apprehension of the field in question in ways that make maximally intelligible *sense*. (Of course, *realization* of that goal for Kant would require the introduction of further structuring – not only intellectual but also, as I mention briefly below, by way of the introduction of instances of intuiting *within* the "original" instance of intuiting, as *sub*-fields for eventual intellectual structuring.)

[48] See footnote 43.

limit or frame is compatible with that pure form. Thus, on the level to which we have so far directed our attention – which might be called that of "original" intuiting or "original" transcendence, that is, the level of consciousness considered apart from the incorporation of intuitings *within* an original intuiting, or of instances of transcendence *within* an original transcendence – it would seem that what we should say for Kant is simply that what is apprehended has the form of a *spatial* or a *temporal* field filled with at least apparent *sensory quality*, the latter in turn then subject to discrimination precisely presupposing (at least in principle conceptualizable) instances of "incorporated" intuiting. For Sartre, the status of spatial form might seem significantly different, inasmuch as he regards space, not as an *object* of any sort of (even "pure") intuiting, but as a mere correlate of the possibility of a certain sort of figure-ground *shift* in the manner in which consciousness is object-directed (254ff/220ff). These matters are discussed by Sartre in the chapter titled "Transcendence", a chapter in which – noting a so far misleading aspect of his presentation – he in fact emphasizes, in the section titled "Determination as Negation" (249ff/216ff), the establishment of "nihilations" within the frame of an all-encompassing *original* nihilation, as the condition for the emergence of any "thises":

> To *what* being is the for-itself presence? ... the question has meaning only if it is posited in a world. ... since it is the presence of the for-itself which brings there *to be* (*fait qu'*il y a) a "this" rather than a "that." Our examples, however, have shown us a for-itself denying concretely that it is a *particular* being. This situation arises from the fact that we described the relation of knowledge before bringing to light its structure of negativity. In this sense, by the very fact that it was revealed in examples, that negativity was already secondary. Negativity as original transcendence is not determined in terms of a *this*; it causes a *this* to exist (*fati qu'un* ceci *existe*). (249/216)
>
> The "this" always appears on a ground; that is, on the undifferentiated totality of being inasmuch as the For – itself is the radical and syncretic negation of it. ... But the appearance of the "this" ... is the correlate of the appearance of my own concrete negation on the syncretic ground of a radical negation. (252/218)

As I have argued elsewhere, at least implicit in Kant too is recognition of the need for an incorporation of intuitings *within* intuitings, up to the point of an "original" all-encompassing intuiting, as a condition for

determinations regarding objects within whatever sort of "field" is given by the latter.[49]

Quite apart from the complication introduced by this last point, the standard picture of Kant's view of space as "form of intuition", as adumbrated above, is arguably overly simplistic.[50] And of course much might be said about differences regarding the status of time in the two thinkers. But what is of relevance to our own concern lies elsewhere. It is simply that, given the ontological status of facticity in Sartre, and its bearing on a correlative field for determinations regarding "thises", the field in question must always have, a priori as it were, the meaning of a *real world* for Sartre, as the all-encompassing field within which, or at least "on the ground" of which, all such determinations need to be made. That is, it must always *itself* have the meaning of "being-in-itself". For Kant, by contrast, whatever determinations are made with respect to whatever might be discriminated *within* it, the given field is *nothing* in-itself: it is, as one might put it, a mere "intentional object (or proto-object)".

In any case, all of this has been intended as preparation for a point about ontological multivalence, that is, for the notion that the very same appearances "determined" as really existent *objects* are also in principle determinable as mere appearances (or at least possible appearances) *of* such objects. This is of course central to transcendental phenomenalism as I have characterized it. According to that view, really existent objects are in some sense "reducible" to series or totalities of appearances. But by contrast with other varieties of phenomenalism, such a reduction requires the equal legitimacy of regarding all those appearances, taken singly, precisely *as* the very object reducible to them. Obviously, there is no mystery in general as to the possibility of one and the same item being regarded in two such radically different ways. But if those items are anything "in themselves", then it seems that at least one of the two radically different regards must simply be mistaken. On the other hand, if they are not anything in themselves, then it is difficult to see how there could *be* any "items" available for such multiple regard in the first place.

---

[49] Aquila (2003: esp. 243ff). I have also developed the point in "Thoughts without Intuition are Blind," delivered at the NYU Conference on Modern Philosophy (November 8, 2013). (A version of this presentation will likely appear in print, perhaps with a different title, in 2015.) For an emphasis on the figure-ground structure in Sartre, cf. Morris (2008: e.g. 63, 188ff).

[50] It is arguable, in particular, that the bare "form" of intuiting yields a field that is at most susceptible to determination *as* one whose form is that of space and time. Cf. Longuenesse (1998[1993]: 215ff), Waxman (2014: 143ff).

As I have suggested, the solution needs to lie in the (as we might put it) "a priori form" of instances of consciousness as instances of *directedness*, where the latter is given independently of any specific *content* of directedness. The price to pay for this must of course be as we have seen: of itself, the a priori form of consciousness in question gives at most a so far *undetermined field* within which determinations remain to be made. Presumably (and as explicitly recognized by Sartre and, as I have claimed, at least implicit in Kant), such determinations in turn require (1) minimally, the incorporation of instances of consciousness of the very *same* form, within an all-encompassing instance of that form (Kantian intuitings within an all-encompassing intuiting, Sartrian "concrete negations" within an original *radical* negation), together with (2) whatever further conditions are (putting the point for simplicity at least in terms that would be favoured by Kant) required in order for it to be the case that those incorporated instances of consciousness are characterized by way of this or that instance of *conceptual content*. We will not consider here how Sartre and Kant might differ with respect to (2). Minimally, emphasis on the "conceptual" might seem too intellectualistic in Sartrean terms. But then, of course, the notion of "conceptual content" might be taken in any number of ways. In any case, the point for our purposes is simply this: by virtue of the a priori form of directedness in question, the entertaining of alternative "determinations", by way of *given instances* of such directedness, within the frame of an all-encompassing field for such determinations, is perfectly reasonably regarded (at least within the terms of anything reasonably regarded as "phenomenalistic") as the entertaining of alternatives with respect to a given "appearance". To this extent, there is nothing ontologically anomalous in the suggestion that the very *same* appearances are alternatively determinable either as actual *objects* or as series of appearances to which such objects are "reducible". In any case, as I have argued, there is good reason to think that it is in the latter way that both Kant and Sartre are thinking.

4 Conclusion

Here is how I have characterized "transcendental phenomenalism" in this chapter:

> *Transcendental Phenomenalism*. Any judgement of an appearance *A*, to the effect that it is a phenomenon (a real existent), is a judgement to the effect that *A* is, from one point of view, a phenomenon and also,

> from another point of view, an appearance within an infinite series of appearances *of* the very phenomenon in question.

I have argued that both Kant and Sartre are proponents of this sort of phenomenalism, and I have shown how to accommodate what might seem incoherent in it in terms of a perhaps surprising further similarity between them, namely, between the status of pure form of "intuiting" in Kant and of Sartrean transcendence of facticity. I have also emphasized a crucial difference: for Sartre but not correspondingly for Kant, by virtue of the pure form of transcendence of facticity, all determinations regarding appearances are necessarily made in some way "within", or "on" or "against", the background of (what is at least *thereby* for consciousness) a world "given" with the meaning of being-in-itself. I have also argued that, contrary to the perhaps apparent import of a number of Sartrean statements, the only sense in which an appearance *as such* has any sort of being-in-itself for Sartre, or is "in itself alone (*à elle toute seule*) a *transcendent being*" (23/27), is two-fold: any appearance is as such (1) at least *possibly* a really existent object (e.g. a table or a package of tobacco) and (2) something for consciousness only "on" or "against" such a transcendent background.

We have also seen that Sartre takes it to be part of the apprehension of appearances *as real existents* that they are apprehended as something in themselves. They are not simply apprehended on or against a *background* whose meaning includes that of being-in-itself; they are apprehended as *themselves* being in that very same sense. This is of course not part of transcendental phenomenalism as such, as I have characterized it, but of the particular variety of it espoused by Sartre. In any case, unlike the sort of "two-aspect" view often attributed to Kant, the ascription of being-in-itself to real existents is presumably not meant to amount for Sartre to regarding them as "things in themselves" in any sort of "transcendental" sense. That is, it seems clear that, for Sartre, to take chairs and packages of tobacco to be in themselves is part of a wholly *empirical* way of regarding appearances as real existents. And we have already seen why Sartre takes this to be so: the closest that a mere series or totality of appearances could come to doing justice to our sense of empirical reality is by way of an ultimately unsatisfying "appeal to infinity".

But then we might ask: why be a "phenomenalist" *at all*? Why not simply make do with our ability to take appearances not simply as appearances but as something "in themselves"? But here too it seems the answer is evident for Sartre: *apart* from an appeal to a series or totality of appearances of which a given appearance is, appropriately, a "member",

but still insofar as it is regarded as a real existent, any appearance to which one attributes being-in-itself would be absolutely *characterless*. It simply would not be for consciousness either a table or a package of tobacco or anything else. That is presumably precisely the point of Sartre's endorsement of at least some sort of "phenomenalism" at the very beginning of "The Pursuit of Being". In any case, there seems to me nothing anomalous, and certainly nothing incoherent, in the following more specific form of transcendental phenomenalism:

> *Transcendental Phenomenalism Sartrean Style*. Any judgement of an appearance *A*, to the effect that it is a phenomenon (a real existent), is a judgement to the effect that *A* is, from one point of view, *something that exists in-itself* and also, from another point of view, an appearance within an infinite series of appearances *of* the very phenomenon in question.

What then of Kant's position?

Here I simply content myself with observing that the position developed in this chapter shows that the number of options available to commentators are at least in principle significantly greater than normally supposed.

(1) (a) *Judgements of empirical reality are sufficiently secured by way of an "appeal to infinity"* with regard to those series of appearances to which the objects in question are "reducible". But (b) those very objects can (or must) *also* be regarded as something in-itself from a transcendental point of view.

(2) (a) [same as (1a)]. But (b) those very objects cannot be regarded as being something in themselves from a transcendental point of view. (Thus any "things in themselves" in a "transcendental" sense must be distinct things, and so part of a "different world".)

And then – all of the following compatible with (3) *Judgements of empirical reality are not sufficiently secured by way of an "appeal to infinity"* with regard to those series of appearances to which the objects in question are "reducible"; appeal to some sort of being-in-itself is also required.

(3a) (i) [same as (3)]. But (ii) appeal to the required being-in-itself amounts to regarding the objects in question as being something in themselves from a transcendental point of view.

(3b) (i) [same as (3)]. But (ii) appeal to the required being-in-itself does not amount to regarding the objects in question as being something in themselves from a transcendental point of view. (The view that I have attributed to Sartre.)

(3c) (i) [same as (3b)]. But (ii) those very objects can (or must) *also* be regarded as being something in themselves from a transcendental point of view. (A kind of two-fold "two-aspect" view.)

(3d) (i) [same as (3b)]. But (ii) those very objects cannot be regarded as being something in-itself from a transcendental point of view. (Thus any "things in themselves" in a "transcendental" sense must be distinct things, and so part of a "different world.")

(3b), again, is the variety of transcendental phenomenalism that I have attributed to Sartre. But all of (1) through (3d) are compatible with transcendental phenomenalism as formulated more generally at the beginning of this section. Furthermore, the very need to consider, at least in principle, such a variety of possible positions is generated by the very condition that renders transcendental phenomenalism possible in the first place, namely, what I have called the ontological "multivalence" of appearances – *together* of course with Kant's own distinction between an "empirical" and a "transcendental" distinction between "things in themselves" and "appearances".

And then finally: What of transcendental *idealism*? With an eye to Kant, there are of course various ways in which one might propose to define that doctrine. I simply conclude by noting that the present approach suggests a natural way of defining transcendental idealism, such that any variety of transcendental *phenomenalism* would at least arguably be only a special case thereof, but not strictly entailed by it:

> *Transcendental Idealism*. Any judgement of an appearance *A*, to the effect that it is a phenomenon (a real existent), is a judgement made with regard to an object whose availability for such judgement is *provided by an instance of the directedness of consciousness that leaves its ontological status altogether open*.

Both Kant and Sartre are transcendental idealists in this sense, and they are also transcendental phenomenalists. What remains debatable is whether they are transcendental phenomenalists because they are in the first instance transcendental idealists in this sense and – rightly or wrongly – take the former to follow from the latter.

References

Aquila, R. E. (1983) *Representational Mind: A Study of Kant's Theory of Knowledge*. Bloomington: Indiana University Press.

—— (1989) *Matter in Mind: A Study of Kant's Transcendental Deduction*. Bloomington: Indiana University Press.

—— (1998) "Sartre's Other and the Field of Consciousness: A 'Husserlian' Reading", *European Journal of Philosophy,* 6: 253–76.

—— (2003) "Hans Vaihinger and Some Recent Intentionalist Readings of Kant", *Journal of the History of Philosophy,* 41: 231–50.

Barnes, H. E. (1992) "Sartre's Ontology: The Revealing and Making of Being", in *The Cambridge Companion to Sartre*. Ed. C. Howells. Cambridge: Cambridge University Press: 13–38.

Carriero, J. (2009) *Between Two Worlds: A Reading of Descartes's* Meditations. Princeton: Princeton University Press.

Descartes, R. (1984) *The Philosophical Writings of Descartes*. Vol. II. Ed. and Tr. J. Cottingham, R. Stoothoff and D. Murdoch. Cambridge: Cambridge University Press.

Gardner, S. (2009) *Sartre's Being and Nothingness: A Reader's Guide*. New York and London: Continuum Books.

Holmes, R. (1984), "Being-In-Itself Revisited", *Dialogue,* 23: 397–406.

Jeanson, F. (1980[1965]) *Sartre and the Problem of Morality*. Tr. R. V. Stone. Bloomington: Indiana University Press.

Kant, I. (1902-) *Gesammelte Schriften*. Ed. Königlich Preussischen Akademie der Wissenschaften. Berlin: Walter de Gruyter and predecessors.

—— (1992[1800]) *Lectures on Logic*. Tr. and ed. J. M. Young. Cambridge: Cambridge University Press.

—— (1998[1781/1787]) *Critique of Pure Reason*. Tr. P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

—— (2002[1783]) "Prolegomena to Any Future Metaphysics". Tr. Gary Hatfield, in *Theoretical Philosophy after 1781*. Ed. H. Allison and P. Heath. Cambridge: Cambridge University Press, 2002.

—— (2005) *Notes and Fragments*. Ed. P. Guyer. Tr. C. Bowman, P. Guyer and F. Rauscher. Cambridge: Cambridge University Press.

Longuenesse, B. (1998[1993]) *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Tr. C. T. Wolfe. Princeton: Princeton University Press.

McCulloch, G. (1994) *Using Sartre: An Analytical Introduction to Early Sartrean Themes*. London: Routledge.

McDowell, J. (2009) *Having the World in View: Essays on Kant, Hegel, and Sellars*. Harvard: Harvard University Press.

Morris, K. J. (2008) *Sartre*. Oxford: Blackwell.

Morris, P. S. (1976) *Sartre's Concept of a Person: An Analytic Approach*. Amherst: University of Massachusetts Press.

Naragon, S. (1990) "Kant on Descartes and the Brutes", *Kant Studien,* 81: 1–23.

Sartre, J.-P. (1943) *L'être et le néant: Essai d'ontologie phénoménologique*. Paris: Gallimard.

—— (1966[1943]) *Being and Nothingness: A Phenomenological Essay on Ontology*. Tr. with an introduction by Hazel E. Barnes. New York: Washington Square Press (originally New York: Philosophical Library, 1956).
—— (1976[1943]) *L'être et le néant: Essai d'ontologie phénoménologique*. Ed. A. Elkaïm-Sartre. Paris: Gallimard.
Sellars, W. (1968) *Science and Metaphysics: Variations on Kantian Themes*. New York: Humanities Press.
Waxman, W. (2014) *Kant's Anatomy of the Intelligent Mind*. Oxford: Oxford University Press.

Index