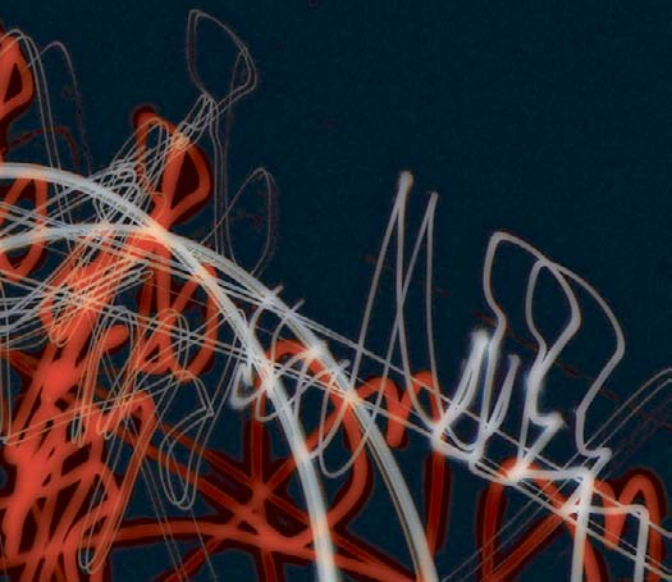
OXFORD

mentalactions

EDITED BY
LUCY O'BRIEN AND
MATTHEW SOTERIOU



MENTAL ACTIONS

This page intentionally left blank

Mental Actions

Edited by
LUCY O'BRIEN AND MATTHEW SOTERIOU

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2009

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Mental actions / edited by Lucy O'Brien and Matthew Soteriou.
p. cm.
Includes index.
ISBN 978–0–19–922598–9 (alk. paper)
1. Act (Philosophy) 2. Agent (Philosophy) I. O'Brien, Lucy, 1964–
II. Soteriou, Matthew.
B105.A35M47 2009
128'.4—dc22 2009004577
Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
MPG Biddles Standard Ltd, King's Lynn, Norfolk
ISBN 978–0–19–922598–9
1 3 5 7 9 10 8 6 4 2

Preface

In 2005 the Institute of Philosophy in London hosted the conference 'Mental Action' at which a sub-set of the papers in this volume were presented. That conference gave us the idea of co-editing a volume on mental action. We thought that, despite the fact that it should be at the centre of philosophy of mind, the topic of mental action had been somewhat overlooked. We proposed to edit a volume that offered something of a corrective to that state of affairs. We approached a number of philosophers who we knew had an interest in actions and agency, to turn their attention to focus on mental actions in particular. We aimed neither for comprehensiveness, nor neutrality. Rather, we approached philosophers that, in our view, were likely to have something worthwhile and substantial to offer on the topic, giving them free reign to determine the focus and scope of their papers. The papers that came in seem to us to constitute an excellent beginning to the process of mining what we hope will be a rich vein in the philosophy of mind and action.

Our thanks are due to a number of people. Most obviously, we want to thank the other contributors to this volume. Thanks also to Tim Crane for organizing and hosting the original conference. Many thanks to Peter Momtchiloff for responding so positively to the proposal for this volume, and to the anonymous referees for OUP for their judgement and hard work. We also want to thank Jane Robson for her efficiency and astuteness as copy-editor, and Hong Yu Wong for excellent work on the index.

Lucy O'Brien and Matt Soteriou

London

Special Acknowledgements

Thanks of a unusually high order are due to my co-editor Matt Soteriou. When family illness meant that I was unable to do my full share as co-editor he swiftly took on the bulk of the work, and patiently waited for me to do my share. Most importantly he agreed to write the introduction for the volume. I am enormously grateful to him for that. The volume would have been much delayed without his willingness to step in. I also want to record my gratitude to my late husband, Mark Sacks. He helped me with the volume, as he helped me with everything, in uncountable ways.

L.O'B.

London

Contents

List of Contributors

Thomas Crowther is Lecturer in Philosophy at Heythrop College, University of London. His research focuses primarily on philosophy of mind and epistemology.

Fabian Dorsch is a Lecturer, and Swiss National Science Foundation (SNF) research fellow, at the University of Fribourg. He is author of *The Unity of Imagining* (Ontos 2009). His research focuses on topics in the philosophy of mind, epistemology, and value theory, notably on the relationship between consciousness and normativity.

John Gibbons is an Associate Professor of Philosophy at the University of Nebraska-Lincoln. He works on the philosophy of mind, epistemology, and action theory with a particular emphasis on reasons and reasoning.

Pamela Hieronymi is an Associate Professor in the Philosophy Department at the University of California, Los Angeles. She specializes in moral psychology, with particular interest in agency and responsibility.

Alfred Mele is the William H. and Lucyle T. Werkmeister Professor of Philosophy at Florida State University. He is the author of *Irrationality* (OUP 1987), *Springs of Action* (OUP 1992), *Autonomous Agents* (OUP 1995), *Self-Deception Unmasked* (Princeton 2001), *Motivation and Agency* (OUP 2003), *Free Will and Luck* (OUP 2006), and *Effective Intentions* (OUP 2009).

Lucy O'Brien is Reader in Philosophy at University College London. She is the author of *Self-Knowing Agents* (OUP 2007). She works on the philosophy of mind and action, focusing in particular on self-consciousness and self-knowledge.

Brian O'Shaughnessy is at King's College London. He is the author of many articles in the philosophy of mind and action, of *The Will: Volumes I and II* (CUP 1980; 2nd edn., CUP 2008), and of *Consciousness and the World* (OUP 2000).

David Owens is Professor of Philosophy at the University of Sheffield. He is the author of *Reason Without Freedom* (Routledge 2000) and *Causes and Coincidences* (Cambridge 1992) and has published papers in epistemology, moral psychology, and ethics.

Christopher Peacocke is Professor of Philosophy at Columbia University in the City of New York, and Richard Wollheim Professor of Philosophy at University College, London. He is the author of *A Study of Concepts* (MIT 1992), *Being Known* (OUP 1999), *The Realm of Reason* (OUP 2004) and *Truly Understood* (OUP 2008). He continues to work in metaphysics, the theory of intentional content, and on the nature of conscious states.

Thomas Pink is Reader in Philosophy at King's College, London. He works on ethics, metaphysics, philosophy of mind and action, philosophy of law, and on the history of philosophy. He is the author of *The Psychology of Freedom* (CUP 1996) and *The Ethics of Action* (forthcoming OUP).

Joëlle Proust is a Director of Research at CNRS at Institut Jean-Nicod (École des Hautes Études en Sciences Sociales et École Normale Supérieure, Paris). She is the author of *Questions of Form* (Minnesota 1989), *Comment l'esprit vient aux bêtes* (Gallimard 1997), *Les Animaux pensent-ils?* (Bayard 2003) and *La Nature de la volonté* (Gallimard 2005). She currently works on the forms of self-awareness involved in bodily and mental agency.

Matthew Soteriou is an Associate Professor of Philosophy at the University of Warwick. His main research interests are in the philosophy of mind and action, and epistemology.

This page intentionally left blank

1

Introduction

Matthew Soteriou

Discussions of action typically tend only to mention mental action in passing, if at all, and more generally, the topic of mental action has received relatively little attention in work in the philosophy of mind. Our aim in bringing together this collection of previously unpublished papers on mental action is to offer something towards a corrective to that state of affairs. What emerges from this work is that there are various ways in which a focus on mental action can contribute to our understanding of a range of different issues and themes—in the philosophy of action, the philosophy of mind, and epistemology. This introduction highlights some of the ways in which the contributions to this volume connect with some of these themes, and with each other.

When turning one's attention to the topic of mental action a natural starting point is to focus on the question of the scope of mental agency. Just which aspects of our mental lives should be regarded as mental actions? This issue is discussed in the first section. In the second section I look at some of the ways in which turning one's attention to mental action in particular might prompt one to retrace and reassess some of the steps taken in general accounts of action arrived at through consideration of the bodily case. Here I also discuss some of the different structures mental agency may take, and whether any such structures are distinctive of mental, as opposed to bodily, action. Finally, in the third section I discuss some of the ways in which a consideration of mental action can illuminate other issues in the philosophy of mind and epistemology. In particular I highlight some of the connections that have been forged between accounts of mental agency and debates about self-knowledge, and the ontology and phenomenology of mind.

I

In commenting on the relative neglect of mental action in treatments of action in general, Michael Stocker once remarked that he found this oversight particularly

strange coming from philosophers, whose professional life is so caught up in mental activity.[1] On the one hand, struck by this thought, one might be tempted to offer some kind of diagnosis of this situation, perhaps by pointing to some background philosophical prejudice or preoccupation that has exerted its influence over action theory. For example, one might speculate that it is perhaps symptomatic of a philosophical preoccupation with the question of the relation between mind and body (or more generally, the mental and the physical). This preoccupation might partly explain the focus on bodily action, given that the agency we exercise over our bodies is often viewed as a central nexus of mind–body interaction, and central also to an understanding of embodiment. On the other hand, one might just regard this relative neglect as a natural reflection of our pre-philosophical intuitions concerning, and interest in, action. For example, perhaps it simply has something to do with the associations we make between the notions of inactivity and immobility, perhaps itself a reflection of our tendency to think that, as O'Shaughnessy remarks, it is what we end up doing in the 'public physical world' that really matters to us.[2]

Could the most straightforward explanation of the relative neglect of the topic be the fact that there is relatively little in our mental lives that is genuinely agential? One striking feature of some of the recent work there has been on mental action is the extent to which opinions appear to diverge on the basic question of the scope of our mental agency. For example, while Peacocke holds that judgings and decidings, as well as acceptings, attendings to something or other, calculatings, reasonings, and tryings can all be instances of mental action, Galen Strawson has argued that, although there is such a thing as mental action, most of our thoughts, including our decisions, 'just happen' and 'action and intention need have little or nothing to do with their occurrence.'[3]

Such disagreement over the issue of the scope of our mental agency can, in turn, lead to very different positions on the question of the significance of the role of mental action in our conscious mental lives. According to Strawson, a reasoning, thinking, judging, self-conscious creature need not be an agent at all, for such mental activity, he suggests, need not involve any mental action. So Strawson suggests that there is no incoherence in the idea of what he calls a 'Pure Observer': 'a motionless, cognitively well-equipped, highly receptive, self-conscious, rational, subtle creature that is well-informed about its surroundings and has, perhaps, a full and vivid sense of itself as an observer although it has no capacity for any sort of intentional action, nor even any conception of the possibility of intentional action.'[4]

[1] Stocker 1982.

[2] O'Shaughnessy 2000: 101. O'Shaughnessy writes, 'Think how one might ruminate for days in making a decision that issues in a simple but momentous act of signing one's name. Then it is the outcome in the *public physical world* that ultimately matters to one's life' (ibid. 101).

[3] See Peacocke 2007; Strawson 2003: 228–9.

[4] Strawson 2003: 228.

With this claim we can contrast the very different view proposed by Brian O'Shaughnessy in *Consciousness and the World*. One of O'Shaughnessy's central concerns there is, as he puts it, 'to discover the contribution made to consciousness, not by the bodily will (for we can be fully conscious though supine in a hammock, and even if totally paralysed), but by the mental will', which, O'Shaughnessy claims, 'cannot in the conscious be analogously incapacitated' (2000: 226). In stark contrast to Strawson's suggestion of the possibility of a 'Pure Observer', O'Shaughnessy suggests that with the imposition of what he calls a 'will freeze' upon conscious experience, 'one cannot but replace the prevailing state of consciousness, waking, with another state of consciousness, perhaps sleep' (ibid. 229). Mental action plays a crucial role in O'Shaughnessy's account of the state of wakeful consciousness in the self-conscious. According to O'Shaughnessy, 'The mind of one who is conscious is necessarily a mind actively governing the movement of its own attention and thinking processes' (ibid. 89). 'In the final analysis it is because thinking is active and thinking is essential to consciousness that mental action is a necessary condition of consciousness' (ibid. 264).

Leaving aside for now the question of whether mental action is a *necessary* condition of consciousness in the self-conscious, it certainly seems intuitive to think that the 'general direction taken by our thoughts and attention' is something that *can* be up to us. Even if we are not always making choices about what to think about when we are awake, we do seem capable of making such choices. But where should we locate the mental actions we perform when we engage in such directed thinking?

Consideration of the question of what can figure in the content of one's intentions in such directed thinking may lead one to think that the role of mental action can, as Strawson puts it, 'at best be indirect'. According to one line of thought, thinking about something involves the occurrence of mental acts individuated, in part, by their propositional contents, and these mental acts can be mental actions only if the particular contents that individuate them are ones that one intends to think. However, in the case of many such mental acts it seems that the content of the mental act cannot figure in the content of one's prior intention.

Strawson has argued that no thinking of a particular thought-content is ever an action. According to Strawson, 'Mental action in thinking is restricted to the fostering of conditions hospitable to contents' coming to mind'.[5] This is because one's thinking of the particular content one thinks can only amount to an action if the content thought is already there, 'available for consideration and adoption for intentional production', in which case, 'it must already have "just come" at some previous time in order to be so available'.[6] Here we might contrast thinking with talking. For example, although some have suggested that judging

[5] Ibid. 234. [6] Ibid. 235.

should be understood as the interiorization of the act of asserting, one might think that the analogies between these acts break down when we consider the role that intention can play with respect to each. Having judged that *p*, one might choose to assert that *p*. Having formed the intention to Φ, one might choose to express that intention. But the mental acts of judging that *p* and deciding to Φ cannot themselves be intended. One may be able to decide to determine (or attempt to determine) *whether p*, and one may be able to decide to decide (or attempt to decide) *whether* to Φ, but one does not then decide to judge *that p*, and one does not decide to decide to Φ.

On this latter point there is wide agreement. In saying that the awake self-conscious subject takes active charge of his own thoughts and attendings, O'Shaughnessy remarks, 'I do not mean actively determines their *content*, which would be at once omnipotent, barren, self-refuting and logically impossible' (2000: 89). And Peacocke concedes that 'when you think a particular thought, there is of course no intention in advance to think that particular thought'.[7] However, for both Peacocke and O'Shaughnessy this still leaves intention with a significant role to play in directed thinking. For Peacocke, directed, as opposed to idle, thought involves 'the intention to think a thought which stands in a certain relation to other thoughts or contents'.[8] Similarly, O'Shaughnessy claims that the intentions involved when one is engaged in such activity select 'the content of the governing enterprise', 'stir one's mental machinery' and constrain, 'under definite description', the advance of one's thinking.[9]

Here we can compare the position taken by Mele in Chapter 2. Mele agrees with Strawson that there is a natural reading of the notion of 'coming to entertain a content' according to which it itself is never an action. However, he argues that even if coming to entertain a content is never an intentional action, that leaves plenty of room for related intentional mental actions, for when we ask why a particular instance of this happens to someone, we may find that mental actions are an important part of the answer.

Mele argues that we need to be careful here to distinguish 'trying to Φ' from 'trying to bring it about that one Φs'. When one tries to bring it about that one Φs, one's Φ-ing can, but, crucially, need not be, an action. For example, falling asleep is not an action, so one cannot, strictly speaking, try to fall asleep, but one can try to bring it about that one falls asleep. Applying this distinction to the domain of mental action, Mele suggests that in the case of many Φ-ings that are not mental actions, the occurrence of the Φ-ing may be explained, in part, by the fact that the subject of the Φ-ing has engaged in the mental actions of trying to bring about that she Φs and bringing it about that she Φs. For example, Mele argues that, although one cannot, strictly speaking, try to remember something, one can try to bring it about that one remembers.

[7] Peacocke 1999: 209. [8] Ibid. 210. [9] See O'Shaughnessy 2000: 89 and 221.

In a similar vein Dorsch (Chapter 3) identifies, within the mental domain, a category of what he calls 'mediated' agency, which involves the instrumental reliance on epistemic or merely causal processes and their passive effects. In the case of 'mediated' agency, we 'trigger' some process (epistemic or merely causal) with some goal in mind, but recognize, and instrumentally rely on, the capacity of such a process to lead, *by itself*, to the desired outcome. According to Dorsch, remembering and judging involve this form of 'mediated' agency, as opposed to supposition and imagining (e.g. visualization), which, according to Dorsch, involve non-mediated, 'straightforward' agency.

For Peacocke, the mental act of judging does not involve a 'mediated' form of agency—the role of mental action in judging is not 'indirect'—for the act of judging is itself a mental action. As we have seen, this is not because Peacocke disagrees with his opponents over whether it is possible to form the prior intention to judge that *p*. It is, rather, because he is operating with a rather different conception of what it is for an event to be an action. According to Peacocke, for a mental event to be a mental action, it must consist of an event which either is, or constitutively involves, a trying, and Peacocke has argued that tryings should be distinguished from prior intentions.

What the requirements are for something to be a genuine action is, of course, itself a contested matter, and so given that there is a range of different views on the criteria for something's being an action, it should come as no surprise that there should exist disagreement as to which mental phenomena count as mental actions. However, while one's views on the scope of mental action may simply be driven by one's prior account of the general criteria for action arrived at through consideration of the bodily case, one could of course be influenced in the other direction. If, for example, one's general account of action leaves relatively little room for the genuinely agential in our mental lives, one might regard this as symptomatic of its having taken a significant wrong turn somewhere. Attention to mental action in particular might then prompt one to retrace and reassess the steps taken in standard accounts of action. The next section highlights some of the different suggestions of this kind that are made in the chapters in this volume.

II

Gibbons argues that in order to accommodate the distinction between the active and passive within the mental domain, we need to reassess some of the assumptions that are often made about the role of reason in action. One of Gibbons's concerns is to provide a way of accommodating the distinction between a mental action of, say, deciding on the one hand, and the event of the mere passive acquisition of an intention on the other. Gibbons argues that causation by practical reasons isn't itself sufficient for being an action, as desires and passively acquired intentions, which aren't actions, can be caused by practical

reasons. However, if we add to this the requirement that the reasons causing the action must be about the action we rule out too much, for we thereby exclude the mental actions of deciding to Φ, judging that *p*, and considering that *p*. For example, to take the case of deciding, Gibbons suggests that our reasons for deciding to Φ are first order. They're not about deciding, they're about what the decision is a decision to do.

Gibbons proposes that intentionally Φ-ing is itself a propositional attitude, and he defends the general claim that n-order attitudes rationalize n-order attitudes. Once we regard intentionally Φ-ing as a propositional attitude, we can then regard an ordinary overt action as the conclusion of practical reasoning. This leads to the proposal that something is an action when it's a step in the process of reasoning, where the relevant notion of 'reasoning' is to be distinguished from mere rational causation. According to Gibbons, the kind of reasoning in question involves having higher-order normative thoughts about your reasons, where your reasons for acting are not the higher-order normative thoughts, but the lower-order propositional attitudes those higher-order normative thoughts concern. This understanding of action suggests a way of circumventing the objection that one's inability to intend to judge that *p* or decide to Φ prevents such mental acts from being mental actions, for it allows us to regard a deciding or judging as an action, insofar as it is a step in the process of reasoning.

Pink (Chapter 5) also argues that certain assumptions commonly made in standard accounts of action should be rejected. However, his proposal and the motivation for it are rather different. Pink argues that the offending assumptions lead to a view of action that doesn't adequately accommodate a notion of self-determination that can capture our intuitions about moral responsibility. He argues that if we are to provide an account of action that makes room for the right account of what we can be held morally responsible for, we will need to give up some of the assumptions that act as an obstacle to the view that our *decisions* are genuine actions, for according to Pink the right account of moral responsibility should accommodate the idea that free agency begins at the will, and is exercised in and through what we decide and intend.

This leads Pink to reject the view of action that he labels the 'voluntariness-based model of action', according to which actions are performed on the basis of some prior motivating pro-attitude towards performing the action, such as a prior intention or decision to perform it. On such a view the motivating pro-attitude—one's prior intention/decision—is itself passive, because decisions cannot be taken just on the basis of prior decisions or desires so to decide. Here we find again the line of thought expressed by Gibbons: if anything moves one to decide on a particular action it is not any prior desire or decision on one's own part to take it, but some reason for acting as decided. Pink's own view is that intentional action is not any expression of prior motivation, but rather a special mode of exercising rationality in its own right—one distinct from and independent of any other. According to Pink's 'practical reason-based

theory of action', what distinguishes actions from other events or changes is goal-directedness, where this goal-directedness involves not simply an event's serving some end, but the event's being motivated by that end. And he argues that this practical mode of exercising rationality is not to be found in voluntary action alone, but is also found in non-voluntary intention-formation. In this way Pink allows for freedom of agency to characterize not just the voluntary action based on a prior decision or intention to act, but also the decision or intention itself.

David Owens is also of the view that if one's theory of action fails to accommodate the agency we exercise over certain aspects of mind, then it won't be able to accommodate adequately the *free* agency we exercise over our bodies either. According to Owens, if we are to secure the right account of practical freedom we will need to secure the notion that certain relevant mental events involved in practical deliberation count as mental actions, and for Owens this includes one's practical judgements—one's judgements concerning what one should do. Owens is in agreement here with Pink that the source of our freedom of action must be free in just the way that action itself is free, but Owens attempts to show that this constraint is satisfied by practical judgement. He argues that, just as one can control one's action by making a practical judgement, one can control whether one makes the practical judgement by making a practical judgement.

Owens suggests that this should make us suspicious of the idea that practical judgement is a kind of belief, if we agree that we don't control the formation of our beliefs as we control our actions and intentions. Crucial to Owens's account, then, is his defence of the distinction between practical judgement and belief. Owens points to three connotations of our familiar notion of belief. He claims that (i) belief is governed by a norm of truth; (ii) belief motivates action on that belief; and (iii) belief is governed by a norm of knowledge. He goes on to defend the claim that it is this third connotation that distinguishes belief from practical judgement.

Hieronymi (Chapter 7) takes a somewhat different stance on these issues. She acknowledges that that over which we exercise control or agency is typically represented in our intentions, for we expect exercises of agency or control to display 'both a certain kind of voluntariness and, relatedly, a certain kind of "reflective distance" or awareness'. She also acknowledges (in agreement with Gibbons and Pink) that when it comes to forming an intention this form of voluntariness and reflective distance is absent, for one can only form an intention to act for reasons one takes to settle the question of whether to act. However, she also suggests that while the formation and revision of belief and intention does not share the structure of ordinary action, it is not plausibly thought of as a mere mental happening either, and this leads her to suggest that, in order to accommodate the agency we exercise over our own intentions and beliefs, we should acknowledge an *additional* category of agency. This form of agency

she labels *evaluative control*, which she contrasts with the more familiar form of agency, which she labels *managerial* or *manipulative* control.

Hieronymi's suggestion is that regarding beliefs and intentions as attitudes that embody their subject's answer to some question or set of questions allows us to conceptualize more clearly the distinct form of agency we exercise over these aspects of mind. A belief that p embodies a positive answer to the question of whether p, and an intention to ϕ embodies a positive answer to the question of whether to ϕ. She suggests that if an attitude embodies our answer to a question or set of questions, then we will form or revise such an attitude in forming or revising our answers to the relevant question(s). So we thereby have a form of control over these aspects of our minds, because as we form or revise our take on things, we form or revise our attitudes.

Here we can contrast Hieronymi's approach with those proposed by Pink and Gibbons. Pink argues that decisions are mental actions. Gibbons argues that some, but not all, decisions are mental actions. Hieronymi argues that while all intentions are subject to some form of agency, none are mental actions. Hieronymi is unhappy with the suggestion that our decisions are mental actions, as she believes (c*ontra* Pink) that the form agency we exercise over our intentions is unlike the form of agency we exercise over our actions. She also rejects the suggestion that we only exercise agency over those intentions of which we are reflectively aware (*contra* Gibbons). According to Hieronymi, 'reflective awareness' is not necessary for the form of agency (evaluative control) we exercise over our intentions and beliefs.

An assumption that appears to be widely shared is that we should seek a unified account of mental and bodily action.[10] Although Hieronymi suggests that we should recognize a distinctive form of agency that we exercise over our minds—that of evaluative control—her proposal is consistent with the assumption that we should be looking for a unified account of mental and bodily action, for she is not suggesting that the mental phenomena over which we exercise this distinctive form of agency should be regarded as mental *actions*. But are there any reasons to think that this assumption might be challenged? Are there any reasons to think that there might be significant differences between bodily and mental actions that don't simply amount to the claim that the one variety is bodily and the other mental? In Chapter 8, O'Shaughnessy argues that at least in the case of a subvariety of mental action there is some further significant difference to be uncovered.

O'Shaughnessy suggests that in the case of all bodily actions, when we affirm 'A did x', we are affirming that A was the agent of an act that we are entitled to designate as 'the active generation of event x'. The 'movement of the will' and

[10] Peacocke (2007) e.g. is explicit that mental actions and bodily actions are actions in exactly the same sense, and that the differences between them are differences between the bodily and the mental.

the event x, which the action is the active generation of, are non-identical, and O'Shaughnessy suggests that it is this non-identity that creates a gap that allows for the possibility of *trying and failing*, and the omnipresence of *trying*. According to O'Shaughnessy it is not possible to conceive of a bodily action in which we do not have this gap, and so it is not possible to conceive of a bodily action for which the concept of trying fails to apply. He also suggests that there are a number of mental actions for which this notion of trying does have application—e.g. trying to remember a name, trying to call up an image, and trying to concentrate. However, he argues that, despite this, there are other examples of mental action that should lead us to qualify a general rule linking action and trying. For there are examples of mental action in which the will is exercised, but where nothing is produced.

In *Consciousness and the World*, O'Shaughnessy argues that silently talking to oneself is just such an example. In his contribution to this volume he argues that the imagining-of the performing of some bodily action is another example of a mental action of this non-productive kind. In such cases, he claims, there is no room left for the phenomenon of trying or striving or attempting. Although the will is operative, trying is absent. Here O'Shaughnessy is in disagreement with Peacocke's assertion that all action must consist of an event that either is or constitutively involves a trying. O'Shaughnessy holds that the general rule that whenever we act we try to do a deed applies without exception in the case of bodily action, but for a variety of mental action this rule breaks down. According to O'Shaughnessy then, the universal rule that applies to all action is that whenever we act the will is operative, but consideration of certain mental actions reveals that not all willings are tryings.

In defence of the claim that mental actions, unlike bodily actions, are not structurally all of one piece, O'Shaughnessy has also argued that the activity of listening, as exemplary of perceptual attendings, presents a 'wholly original schema for action'.[11] In Chapter 9, Crowther examines and develops O'Shaughnessy's views on listening. One might think that if there are any perceptual activities in which agency plays some role that role must at best be 'indirect' or 'mediated', for as Crowther puts it, a familiar thought is that 'in perception we are passive and at the mercy of our immediate environment'. However, such perceptual activities as looking, watching, and listening appear to involve something agential. Should we understand this form of agency on the model of Dorsch's 'mediated' agency—the triggering of some process upon which we instrumentally rely in order to achieve some desired outcome—e.g. one's seeing or hearing something?

Crowther considers and rejects the proposal that the activity of listening to something has an instrumental structure—involving an aural task that has hearing or aurally noticing as its end. He argues that listening to O is not a process

[11] See O'Shaughnessy 2000: 383–406.

that is necessarily terminated by hearing O, for one generally hears O throughout a period of time during which one listens to it. Crowther's own proposal is that listening to a sound involves listening to the producer of that sound, and listening to the producer of a sound is a process of agentially maintaining aural perceptual contact with that producer with the aim of knowing what sound it is producing. He argues that listening to something, understood as a process of *maintaining* or *sustaining* hearing, is not analysable in instrumental terms, in contrast to listening out for something. According to Crowther, there's no *productive* relationship between listening and hearing. Listening does not produce the occurrence of an event of hearing. Listening entails hearing because listening is the agential preservation of hearing.

One's interest in the topic of mental action may not simply lie in the light it throws on an account of action in general. As well as furthering our understanding of which aspects of our mental lives are genuinely agential, and of the different structures such agency may take, it is plausible to think that a consideration of mental action may connect with, and illuminate, other issues in the philosophy of mind. A suggestion touted earlier was that the relative neglect of mental action might be symptomatic of a philosophical preoccupation with the question of the relation between mind and body. One response to this is to mark a distinction between the mind/body problem (the question of how the workings of the body are related to the workings of the mind) and the agent/mind problem (the question of how the workings of the mind are related to the activity of the agent). Although these two problems might be combined to yield an agent/body problem, it might be thought that the best strategy is to start by pursuing each independently.[12] However, although this division of philosophical labour may help to make room for an interest in mental action for action theorists, it may have the further consequence of leading other philosophers of mind to neglect the relevance of mental action to some of their central concerns. For example, once the question of the relation between the mental and physical comes to dominate our philosophical concerns with consciousness, the danger is that we may overlook the potential significance of mental action to our understanding of consciousness.

As has been touched upon already, Brian O'Shaughnessy has made a distinctive contribution to the case for thinking that mental action should be of central concern to philosophers of mind in their attempts to understand the nature of consciousness. In recent work Peacocke has also suggested both that much conscious thought consists of mental actions, and that an account of the nature of our mental actions, and our knowledge of them, can provide a clarification and explanation of a range of features present in conscious thought, as well as helping us to address various classical philosophical issues about the mental,

[12] I owe this way of summarizing such an approach to David Velleman's seminars on action, previously published on his website (http://homepages.nyu.edu/~dv26/).

self-knowledge, and the first person. In general, then, one might look for various ways in which a focus on mental action might contribute to our understanding of a range of issues and themes: in the philosophy of mind and epistemology, as well as the philosophy of action. In the final section I highlight some of the ways in which the contributions to this volume connect with further issues in the philosophy of mind and epistemology.

III

For O'Shaughnessy, the claim that mental action has a crucial role to play in an account of the state of wakeful consciousness in the self-conscious depends, in part, on the link he sees between 'operations of the mental will' and the form of self-knowledge, and rationality of state, that he claims is distinctive of waking consciousness. One of his suggestions is that there obtains a form of self-knowledge in the conscious, self-conscious subject that does not obtain when such a subject is not awake—e.g. when dreaming. According to O'Shaughnessy, in the case of the awake subject there is a form of self-knowledge concerning the progression of his stream of conscious thought and imagination that is explained, in part, by appeal to his intentions in engaging in agential mental activity. The intentions formed by the agent bind the parts of a train of conscious thought together, and such intentions explain why 'The rationale from moment to moment of the progression is openly accessible to one'.[13] Roughly speaking, the idea appears to be that the awake conscious agent is able to make sense of what is happening in a certain domain of his mental life in so far as he is able to make sense of what he is doing, and he is able to make sense of what he is doing in so far as it was his idea to begin with.

Whether or not one agrees with O'Shaughnessy's proposal that mental action (and the distinctive form of self-knowledge that accompanies it) is a necessary condition of the state of waking consciousness in the self-conscious, it is plausible to think that acknowledging that the perspective one has on one's mental life can be that of an agent may have significant implications for the epistemology of mind. Certain aspects of one's knowledge of one's own mental life will be a matter of knowing that one is doing something and knowing what one is doing.

One's stance on which aspects of mind one knows in this way, and how one knows them, will be affected by one's views on the scope of mental agency and one's views on the epistemology of action in general. Peacocke's account of how we standardly know our own actions appeals to the occurrence of belief-independent events of action-awareness that have a first-personal, present-tensed content of the form 'I am doing such-and-such now'. Action awareness is

[13] O'Shaughnessy 2000: 218–19.

standardly brought about by the event of trying that causes the action that the action awareness represents; and according to Peacocke, the distinctive way in which a subject comes to know of his own actions is by taking such an apparent action awareness at face value.

Peacocke (Chapter 10) argues that this distinctive action awareness exists for mental actions, as well as for bodily actions, and as we have seen, Peacocke holds that one's mental actions include judgings and decidings, as well as calculatings and reasonings. So for Peacocke, the distinctive way in which a subject comes to know that she is judging that *p* is by taking at face value an action awareness with the content 'I am judging that *p*'. Peacocke explains why this way of coming to have beliefs about one's mental actions is a way of gaining knowledge, by appealing to the possession conditions for concepts of mental actions. According to Peacocke, such possession conditions contain clauses about first-person present-tense ascriptions that say that the thinker has reason for making such ascriptions in the presence of suitable apparent action awareness.

Peacocke notes that the account of the structure and underlying explanation he offers of the entitlement relations involved in taking action awareness at face value are parallel to ones that could be supplied for our entitlement to take certain observational contents of perceptual experience at face value. However, he argues that, although action awareness makes available demonstrative ways of thinking of actions given in one's action awareness, it is not perceptual awareness—a subject can have action awareness of something without having any perceptual awareness of it. He defends his account of the role of action awareness in our knowledge of our mental actions against the kinds of objection that have been raised against perceptual models of introspection, and goes on to offer an account of how his action awareness proposal can reconcile externalism about intentional content with privileged self-knowledge.

O'Brien also connects issues concerning mental agency and self-knowledge with debates about the tenability of an externalist account of content. In Chapter 11 O'Brien argues that a certain kind of externalist view can be shown to be particularly problematic when we focus on the role of mental action in our thinking. The externalist view that is her target involves a commitment to the following claims: (*a*) there are demonstrative thought-contents such as 'That glass is heavy' which are object-dependent, so if there is no object of the relevant kind there is no content; (*b*) there is just one type of content that characterizes our thought, so thought is not, for example, constituted out of a pair of dual contents; and (*c*) all thoughts are constituted by contents, so if there is no content there is no thinking. The mental action O'Brien focuses on is supposition. The problem she raises for this kind of externalist account is to explain what is going on when an attempt to make a supposition with a demonstrative content fails due the fact that there is no relevant object available to be demonstratively referred to, and hence no content to suppose.

A familiar claim in the case of bodily action is that it can seem to a subject as though she Φ-ed (e.g. raised her arm) when in fact she failed to do so. The suggestion often made here is that if we take tryings to be antecedents of actions we can explain what is going on in such cases by appealing to the idea that while the subject failed to raise her arm, she nevertheless tried to raise it, and her trying to raise it is what explains why she thinks that she raised it when she did not. However, O'Brien argues that a proposal of this form is not available to the object-dependent externalist when it comes to trying to explain what is going on in a case in which a subject makes a failed attempt to make a supposition with a demonstrative content. In the case of a mental action that involves a failure of content there does not seem to be the possibility of retreating to a trying or intention to explain what is going on. The object-dependent externalist cannot explain what is going on by saying that the agent carried out an action of trying to suppose that *p*, for if there is no possibility of supposing that *p*, due to content failure, then there is no possibility of trying to suppose that *p*, due to content failure. O'Brien considers a number of different responses that the object-dependent externalist might offer in response to this problem and argues that all are, in one way or another, inadequate. In particular, they fail to provide an adequate answer to the ontological question: what is going on in the mental life of the subject when this kind of failure of self-knowledge takes place?

The way in which an account of mental agency and the self-knowledge that accompanies it may relate to ontological concerns is also a theme of Soteriou's chapter. Like Peacocke, Soteriou is concerned to avoid a perceptual model of introspection in accounting for the way in which a subject is aware of the conscious mental acts that constitute the mental action she is engaged in, and he connects this epistemological concern with issues in the ontology of mind. In particular, he addresses the question of the ontological category of the conscious mental acts an agent is aware of when engaged in such directed mental activities as conscious calculation and deliberation.

Soteriou argues that if we are to accommodate the idea that an agent can be aware of *doing* something when consciously thinking, then what the agent is aware of must include the occurrence of mental events, and not just the obtaining of mental states. Such mental events, he suggests, should not be regarded as events that are the mere acquisitions of mental states and nothing more, for then they would be the kinds of events that the agent could only have access to via her access to the states that are acquired, in which case, awareness of such events would not be present-tensed. This suggests that the kinds of events in question must be ones that have temporal extension—events that unfold over time. The problem Soteriou then goes on to address is how to reconcile this line of thought with Geach's arguments for the claim that mental acts like judging lack temporal extension.

The suggested solution involves the claim that we should think of the conscious mental act of judging as involving the occurrence of a phenomenally conscious

act with duration that is the *vehicle* of the mental act of judging, just as, in the case of thinking out loud, the bodily action of one's saying that *p* is the vehicle of one's judging that *p* out loud. According to Soteriou, the phenomenally conscious mental act that unfolds over time and which is the vehicle of one's consciously judging that *p*, should be regarded as an event that manifests one's knowledge of what one is thus doing, which involves regarding the event as manifesting a mental state that plays a particular kind of role in the mental life of the subject.

One of the issues addressed in Soteriou's chapter is whether or not mental actions like calculation and deliberation need involve the occurrence of *phenomenally* conscious events. A further phenomenological concern that a consideration of mental action raises is whether there is a sense of agency in thinking. Peacocke has suggested that mental actions, including judging and deciding, have the phenomenology of doing something, rather than involving the phenomenology of being presented with something as being the case, as in perception, or as something occurring to one, as in unintended imagination. Dorsch, on the other hand, wishes to emphasize the respect in which we experience our judgements as passive.

Recall that, according to Dorsch, remembering and judging involve a form of 'mediated' agency, which involves the instrumental reliance on epistemic or merely causal processes and their passive effects, whereas supposition and imagining involve non-mediated, 'straightforward' agency. Dorsch proposes that we consciously experience the 'straightforward' results of our mental agency as practically motivated—as occurring in *immediate* response to practical reasons—and so as active, whereas we experience our judgements as epistemically motivated. Furthermore, Dorsch argues that we cannot experience a mental episode as both epistemically and practically motivated. Hence, none of our mental episodes can be phenomenally marked for us both as a judgement and as a product of 'straightforward' agency. According to Dorsch, this explains why we cannot form judgements at will: 'we would have to consciously experience the resulting episodes in a way which is not open to us'. He contrasts his phenomenological explanation of why we cannot form judgements at will with attempts to reach the same conclusion by appealing to the notion that judgements (or beliefs) are intrinsically and normatively linked to truth.

In Peacocke's account, there is no commitment to the claim that one experiences one's judgements as practically motivated, but he has suggested that through an action awareness of one's judging that is caused by an event of trying, one is aware, non-perceptually, of that thinking as something that one is doing. He has speculated that it is this awareness of one's own agency in thinking that is missing in the schizophrenic experience of 'thought insertion'—the experience of one's own conscious thought as caused by an external, intervening agent.[14]

[14] See Peacocke 2007.

The suggestion that subjects suffering from this symptom of schizophrenia exhibit a disturbed sense of agency in their thinking is discussed by Proust in Chapter 13. However, Proust makes a rather different proposal about where to locate the proper level at which subjects usually feel agentive in their thinking. According to Proust, the proper level at which you feel agentive in thinking is when you assess (predictively or retrospectively) your capacity to perform a given mental action. She suggests that the mental actions we may choose to engage in can be evaluated and predicted for adequacy, and that part of this activity is performed unconsciously. The empirical hypothesis she considers is that a set of *comparators* allows one to *anticipate* how things normally develop for a given type of mental action (say: a directed remembering, or a planning), and that these anticipations are then compared with actual feedback. The mental agent has immediate subjective, phenomenological access to the comparator's verdict. These 'metacogonitive' epistemic feelings express the degree of subjective epistemic uncertainty or sensed feasibility for a given task. For example, in a 'tip of the tongue' experience, a subject becomes aware both that she is failing to retrieve a memory, and that it is worth trying harder.

According to Proust, these metacognitive feelings are the basis on which the subject can judge (*a*) that she can perform a given mental action, (*b*) that she has reached her goal, or failed to reach it, or (*c*) that she has performed no action at all. Proust suggests that these metacognitive feelings may be a necessary structure for feeling responsible for one's mental actions, in which case an inability either to have them in the normal way, or to use them in controlling rememberings, plannings, decidings, etc., may, in part, be responsible for the disturbed sense of agency involved in the experience of 'thought insertion'.

In these introductory comments I have not tried to offer an exhaustive list of the issues and themes addressed in this volume. There are many more for the reader to explore. Indeed, one striking, and exciting, aspect of the work that has been done on mental action so far, is the sheer variety of issues that a consideration of mental action has the potential to illuminate: key concerns in the metaphysics, epistemology and phenomenology of mind, as well as issues connected with our understanding of action in general.[15]

REFERENCES

O'Shaughnessy, B. (2000) *Consciousness and the World* (Oxford: Oxford University Press).

Peacocke, C. (1999) *Being Known* (Oxford: Oxford University Press).

[15] I'm very grateful to Tom Crowther for comments on an earlier version of this introduction. Special thanks are due to Lucy O'Brien for her detailed comments and invaluable advice.

Peacocke, C. (2007) 'Mental Action and Self-Awareness', in J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind* (Oxford: Blackwell).
Stocker, M. (1982) 'Responsibility Especially for Beliefs', *Mind*, 91/363: 398–417.
Strawson, G. (2003) 'Mental Ballistics or the Involuntariness of Spontaneity', *Proceedings of the Aristotelian Society*, 103/3 (April): 227–56.

2

Mental Action: A Case Study

Alfred Mele

According to an attractive, 'causalist' conception of intentional actions, they are, essentially, events with a suitable causal history featuring pertinent mental events or states or their neural realizers. In 'Agency and Mental Action' (Mele 1997), I argued that intentional mental actions pose no threat to causalism that turns on their being mental (as opposed to overt) actions. (Overt actions are actions that essentially involve peripheral bodily motions.) I have not changed my mind about that, but I do think that I should have been more careful when attributing mental actions to agents. The alleged examples of mental action mentioned in that article are 'solving a chess problem in one's head' (p. 231), 'deliberating about whether to accept a job offer' (p. 231), thinking of seven animal names starting with 'g' (in response to a request to try to do so, pp. 234–5), solving 'a complicated arithmetic problem in [one's] head' (p. 240), deciding to *A* (pp. 240–3), various focusings of attention on tasks, and various mental tryings. The example that troubles me now is the one about animal names.

Consider the following from Mele 1997:

> Take a break from this paper and think of seven kinds of animal having a name starting with 'g.' Try it. What happens? Appropriate animal names start coming to mind: 'giraffe,' 'goat,' 'groundhog'. . . . If you are like me, seven suitable names do not leap quickly to mind. You must keep your attention focused on the task if you are to complete it. Suppose that you so focus your attention and—in the absence of deviant causal chains and other monkey business—your so doing results in your acquiring a conscious mental list of seven suitable names. Then you have performed an action describable as 'thinking of (or mentally identifying) seven kinds of animal having a name starting with "g".' (p. 234)

I now believe that even though the words that appear in quotation marks in the final sentence of this passage may describe an action, they do not describe it *as* an action. I will explain why with the aim of shedding light on the nature of mental action. With the same aim, I will explore the merits of various alternative ascriptions of (intentional) mental actions to a normal agent who succeeds at the animal-names task. In section 4, I apply the results to some recent work on mental action.

1. MENTAL ACTION AND TRYING

Gail completes the animal-names task in about three minutes, with no intermission. In what did her thinking of seven animal names starting with 'g'—her *7*-ing—consist? Wittgenstein asks (1953: 622): 'What is left over if I subtract the fact that my arm went up from the fact that I raised my arm?' One answer is his *trying* to raise his arm, which answer may be supplemented by the claim that his (intentionally) raising his arm is identical with his successful attempt to raise it (Adams and Mele 1992). What is left over if we subtract the fact that seven animal names starting with 'g' came to mind for Gail from the fact that she *7*-ed? Perhaps Gail's trying to *7*; and perhaps Gail's *7*-ing is her successful attempt to *7* (Mele 1997).

But perhaps not. Hugh McCann remarks, instructively, that the word 'trying' 'signifies the business of going about the performance of the action' (1986: 201). Was Gail engaged in the business of going about the performance of an *action* of *7*-ing? The question whether Gail's *7*-ing was an action is linked to the question what Gail tried to do.

Trying plays a central role in my treatment of mental actions in Mele 1997, and it does so in some subsequent work on the topic by others (Buckareff 2005; Proust 2001). Trying to *x*, as I understand it, is making an effort to *x*, however modest that effort may be. When, for example, I intentionally sign my name, I am trying to do that, even if I encounter no special resistance and consequently make no special effort (see Adams and Mele 1992 for support). I am, in McCann's words, engaged in 'the business of going about the performance of the action' of signing my name; and my engagement in that business requires some effort, but not much. In Mele 1997, I made no mention of a distinction between trying to *x* and trying to bring it about that one *x*-s. That distinction bears on the present issue. Perhaps Gail tried to bring it about that she *7*-ed and succeeded in that, but did not, strictly speaking, try to *7* and did not perform an action of *7*-ing.

Consider requests to try to sneeze and to try to believe that you are not reading. If you are anything like me, trying to sneeze is not an option for you, but you can try to bring it about that you sneeze. Sniffing a little ground pepper would work for me. Similarly, trying to believe that you are not reading this is not an option for you, but you can try to bring it about that you believe that you are not reading this. Simply stop reading. Primed by the preceding sentences, you will believe, occurrently, that you are not reading.

Here are two requests with a familiar ring: 'Try to fall asleep'; 'Try to remember what you had for dinner three nights ago.' I suggest that they are convenient shorthand for requests with a more stilted ring: namely, 'Try to bring it about that you fall asleep' and 'Try to bring it about that you remember what you had for dinner three nights ago.' Falling asleep is never an action;

it is something that happens to us. That, I suggest, is why you cannot, strictly speaking, try to fall asleep.[1] But people can try to bring it about that they fall asleep, and they can succeed. Counting sheep works for some people. Others resort to sleeping pills. To say that falling asleep is something that *happens* to us is not to say that we cannot bring it about that it happens.

Remembering that *p*, as I understand the phenomenon, is never an action either; and the same goes for remembering *x* (e.g. a face, an odor). If that is right and if the things that agents can, strictly speaking, try to do include no nonactions, you cannot, strictly speaking, try to remember what you had for dinner three nights ago. But if you are like me, you can try to bring it about that you remember this. I have various memory-priming strategies for doing this. One is to ask myself (silently) what I had for dinner on that day and to keep my attention focused on that question. Normally, the answer does not come to mind straightaway, and I pursue the memory-priming strategy further by asking myself (silently) what other things I did on that day. An alternative strategy is to ask myself (silently) what I had for dinner last night and, if the answer comes to mind, work backward—which requires keeping my attention focused on my task. Often, I supplement the second strategy with the first.

Memory-priming is an interesting phenomenon in its own right, of course. One can even bring it about that one remembers more of one's dreams than one normally does. A relatively simple strategy is to ask oneself (silently or otherwise) the following three questions when one awakes—What did I dream? Was I the main character in my dream? How might I have changed the course of my dream if I had been in control of it?—and to focus one's attention on these questions for a while. After a few consecutive days of this, one tends, as a consequence, to remember more of one's dreams than had been the case.

The application of the distinction between trying to *x* and trying to bring it about that one *x*-s is not limited to cases in which *x* is not an action. Unintentional actions are, of course, actions. Imagine that someone offers Ann $1000 for offending Bob unintentionally next week.[2] A straightforward, successful attempt to offend Bob will not gain Ann the money, as she realizes: such an attempt would amount to intentionally offending him. Nor can she see how to try to offend him unintentionally. However, Ann does have an idea about how she might bring it about that she offends Bob unintentionally. Ann knows that she tends to offend Bob unintentionally when she is extremely busy: when she is preoccupied with her work, for example, she tends, without then realizing it, to speak much more tersely than she ordinarily does to people who phone her at the office; and, when Bob calls, her terse speech tends to offend him. Knowing this, Ann undertakes an engrossing project—writing a paper

[1] Someone who mistakenly thinks that falling asleep is an action might do something that he calls 'trying to fall asleep.'

[2] In this paragraph, I borrow from Mele 1995: 413–4.

on mental action—with the hope that her involvement in it will render her telephone conversation at the office sufficiently terse that, should Bob call (as he frequently does), she will unintentionally offend him. In pursuing this strategy, Ann is trying to bring it about that she offends Bob unintentionally. For the record, her attempt is successful.

The distinction at issue even applies to cases in which *x* is an intentional action. Carl has been procrastinating about painting his kitchen. He tries to bring it about that he paints it tonight by telling his children at breakfast that he intends to paint it tonight. As Carl sees it, once he makes the announcement, his failing to follow through would carry a cost that he is unwilling to pay: his children would think less of him for backing out. His plan is to bring it about that he paints his kitchen tonight by significantly increasing his motivation to paint it tonight. Obviously, in making his announcement at breakfast, Carl is not trying to paint his kitchen tonight; but in making his announcement, he is trying to bring it about that he paints it tonight.

These illustrations of the distinction between trying to *x* and trying to bring it about that one *x*-s provide guidance for someone concerned to ascertain whether, strictly speaking, Gail should be said to try to *7* or instead to try to bring it about that she *7*-s. But more guidance is needed. If, as I suggested, the things that agents can try, strictly speaking, to do include no nonactions, what that restriction amounts to depends on what actions are or, more cautiously, on the operative sense of 'action' in the present context. Attention to trying sheds some light on action in general, and the converse also is true.

2. SENSES OF 'ACTION' AND THE INDIVIDUATION OF ACTIONS

There are broader and narrower uses of 'action' in ordinary English. In a very broad use, there are actions of acids, winds, and waves. For example, acids dissolve things, winds blow things around, and waves push and drag things; and these events count as actions in a broad sense. Philosophers of action tend to have little interest in events that are actions only in this very broad sense. They are primarily interested in intentional actions; but, in this very connection, they are also interested in unintentional actions. Part of what is required for understanding what it is to perform an intentional action is understanding how doing that differs from performing an unintentional action.

Donald Davidson contends that 'action . . . require[s] that what the agent does is intentional under some description' (1980: 50). He also contends that 'a man is the agent of an act if what he does can be described under an aspect that makes it intentional' (p. 46). Putting these remarks together, we get the thesis that *x* is an action if and only if *x* is an intentional action under some

description.[3] Obviously, Davidson is not trying to explicate the broad sense of 'action' that I mentioned.

Readers who are not familiar with Davidson's theory of action individuation may find the preceding paragraph opaque. A sketch of the theory will help. 'I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home' (Davidson 1980: 4). How many actions has the agent, Don, performed? Davidson's *coarse-grained* answer is one action 'of which four descriptions have been given' (1980: 4). One obvious upshot is that the same action may be intentional under some descriptions and unintentional (or not intentional) under others. For example, Don performs an action that is intentional under the descriptions 'flips the switch,' 'turns on the light,' and 'illuminates the room' but unintentional (or not intentional) under the description 'alerts the prowler.'

Unfortunately, this theory does not directly yield a diagnosis of why it is that someone who intentionally brings it about that he sneezes by sniffing pepper does not perform the action of sneezing. During lunch, Edna bets Fred $20 that he cannot sneeze within thirty seconds. Fred, a quick thinker, says 'You're on,' sniffs some pepper, and sneezes—all within thirty seconds. 'What [Fred] does' is intentional under various descriptions: for example, 'brings it about that he sneezes' and 'sniffs pepper.' Even so, his sneezing is not an action in the sense of 'action' that Davidson tried to elucidate. (Maybe it is an action in a broad sense of the term in which acids act.) It may be suggested that the answer is to be found by focusing on the brief span of time during which Fred sneezes. However, he is trying to cover his nose and mouth with his left hand at that time (so as to avoid offending Edna), and he succeeds in this. 'What [Fred] does' during that span of time is intentional under various descriptions.

One alternative to Davidson's theory is Alvin Goldman's *fine-grained* theory. It treats actions *A* and *B* as different actions if, in performing them, the agent exemplifies different act-properties (1970).[4] On this view, Don performs at least four actions, since the act-properties at issue are distinct. An agent may exemplify any of these act-properties without exemplifying any of the others. Someone may even turn on a light in a room without illuminating the room: the light may be painted black. A fine-grained theorist can say that Fred's sneezing is not an action because sneezing is not an act-property. (Of course, we would want to be told why it is not an act-property.) But this does not help with related cases. For example, some instances of blinking are actions and some are not. Goldman understands an act-property as 'a property such that at least one of its instances is an act-token' (1970: 16). On this view, some exemplifications of the act-property blinking are actions and some are not.

[3] Davidson expresses the point differently: 'a person is the agent of an event if and only if there is a description of what he did that makes true a sentence that says he did it intentionally' (1980: 46).

[4] For another alternative, the *component* theory, see Ginet 1990: ch. 3.

In this article, I leave both of these competing theories of individuation open. Readers should read such expressions as 'Fred's sniffing pepper' in accordance with their preferred theory of action individuation (and the same goes for the term 'action'). Those who favor the coarse-grained theory should understand this expression as shorthand for 'an action under the description "Fred's sniffing pepper," ' and other readers should take it at face value.

3. GAIL'S INTENTIONAL ACTIONS

Return to Gail and her thinking of seven animal names starting with 'g' (her *7*-ing). How might one argue for the claim that her *7*-ing is not an action?

The following argument uses my claim that remembering is never an action as a premise. The animal names starting with 'g' that Gail thinks of are names that she has learnt. She undertakes a memory-priming project for appropriate words. Furthermore, there is nothing more to her thinking of 'goat,' for example, than her consciously remembering the word; and the same is true of her thinking of each of the other words on her eventual mental list of animal names starting with 'g'. What it is for Gail to think of an animal name starting with 'g' is for such a word to come to mind—for Gail to become conscious of some word of this kind. And her becoming conscious of such a word, in the case at hand, is precisely her consciously remembering the word. Consequently, because remembering is never an action, we should conclude that, regarding each of those words, Gail's thinking of it is not an action. There also is nothing more to Gail's thinking of seven animal names starting with 'g' (her *7*-ing) than her consciously remembering seven such names. So, because remembering is never an action, her remembering seven such names is not an action, in which case her *7*-ing is not an action.

How powerful is this argument? If there is more to Gail's thinking of 'goat'—that event—than her consciously remembering the word, then the argument fails. Now, Gail's thinking of 'goat' has such relational properties as being a thinking of a word that is appropriate to her task, but so does her consciously remembering the word 'goat.' It may be claimed that her thinking of 'goat' includes her recognizing that it is an animal name and that it starts with 'g,' whereas her remembering the word does not. I disagree. Gail's thinking of the word 'goat' is one thing, and her recognizing that it has the features identified is another.

It is possible to extend the series of replies and counter-replies, but some general observations would speed things up. First, a very plausible hypothesis is that Gail's memory for animal names starting with 'g' is primed by her accepting the animal-names task. Second, Gail's thinking of an appropriate 'g'-word should not be confused with anything she does that helps to bring it about that she thinks of that word; for events should not be confused with their causes. (Some such

things that Gail does are discussed shortly.) Third, nor should Gail's thinking *of* a word be confused with, for example, her thinking *that* the word starts with 'g.'

Is there more to Gail's thinking of seven animal names starting with 'g' (her *7*-ing) than her consciously remembering seven such names? As I will explain shortly, there is more to her *bringing it about that* she thinks of seven such names than her consciously remembering seven such names and the former is an action. But I have already explained that bringing it about that one *x*-s should not be confused with *x*-ing and that performing an action of bringing it about that one *x*-s does not entail performing an action of *x*-ing. In the same vein, Gail's *7*-ing should not be confused with the combination of her *7*-ing and things she does that help to bring it about that she *7*-s. The claim that there is nothing more to Gail's *7*-ing than her consciously remembering seven animal names that start with 'g' obviously is consistent with the claim that Gail performs various actions that help to bring it about that she *7*-s.

As a partial consequence of her keeping her attention focused on the task and other mental actions to be discussed shortly, Gail eventually consciously remembers seven animal names that start with 'g'. Her remembering seven such names is not itself an action, and her remembering seven such names is what the argument at issue claims her *7*-ing is. A critic may claim that, although the assertion that Gail thinks of seven animal names starting with 'g' can legitimately be interpreted as the assertion that she remembers seven such names, and although her remembering seven such names is not an action, Gail does perform a mental action that may be *misdescribed* as 'thinking of seven animal names starting with ''g''.' But this is not, in fact, a criticism. Some proponents of the argument at issue have no wish to disagree.

Is the argument part of a merely verbal dispute? Galen Strawson, in a recent article on mental action to be discussed in section 4, is engaged in a dispute of a very similar kind; and his thesis is not that he and his opponents are simply using the same words differently, but that his opponents are wrong. For example, he writes: 'I am directly opposed to Peacocke, . . . who advances from the claim that ''judgements are in fact *actions*'' and the claim that ''to make a judgment is the fundamental way to form a belief'' to the conclusion that coming to form a belief is also standardly a matter of action' (2003: 238; also see Audi 1999).[5] To judge now that either dispute is merely verbal would be hasty.

If Gail's *7*-ing is not an action, is that bad news for mental action? Well, Gail's trying to bring it about that she *7*-s is a mental action—indeed, an intentional mental action. Gail's bringing it about that she *7*-s—her *B7*-ing—also is an action. Now, her *B7*-ing certainly is not a basic action.[6] So it is appropriate to ask *how* she *B7*-s. Her keeping her attention focused on the task (Mele 1997: 234)

[5] Strawson here cites Peacocke 1999: 19–20, 238.

[6] Roughly speaking, *basic* actions differ from non-basic actions in not being performed by way of performing another action.

is important. Gail also silently keeps a running count of the 'g'-names that she has thought of in order to be in a position to know when she has thought of seven; and, as her list grows, she silently repeats to herself the 'g'-names that have occurred to her in order to avoid counting the same name twice. After the first three 'g'-names—'giraffe,' 'goat,' and 'groundhog'—came to mind, Gail felt a bit stuck. She noticed that only mammals were on her mental list, and it occurred to her than she might do well to turn to fish and then insects or birds, if she stalled out on fish. Gail tried to bring it about that she thought of fish names starting with 'g'. The attempt began with her silently asking herself what fish names start with 'g'. Nothing came immediately to mind, but she kept her attention focused on the task. Soon 'goldfish' came to mind, followed by 'guppy.' Then Gail turned to insects and thought of 'gnat.' Her next thought was 'gnu' (which is not surprising, given its similarities in spelling and sound to 'gnat').

In what did Gail's *B7*-ing consist? Here is a short answer: in her successful attempt to bring it about that she *7*-ed. A fuller answer would appeal to a fuller account of trying to *A* than I have given here thus far. A comparison of Gail's case with a case of overt action will help.[7] Young Hal is inscribing on paper what looks like a multiplication problem and the various steps to the solution. He inscribes '111,' '111' again, immediately under the former, and then draws a line under the latter. He inscribes another '111' directly under the line, then another under the previous one, but shifted a bit to the left, then another under it, shifted further to the left. Is Hal multiplying numbers (111×111)? As it happens, *no*. He is trying to draw the leaning tower of Pisa. In virtue of what is it true that he is trying to do the latter and not trying to do arithmetic? A natural answer is that he intends to draw the leaning tower of Pisa (while having no such attitude toward multiplying numbers) and this intention is causally related 'in the right way' (whatever that may be) to his marking the paper as he does.[8]

In Gail's case, an answer of the same form is available to the question in virtue of what it is true that she is trying to *B7*. The answer is that her keeping her attention focused on the task of *B7*-ing, silently keeping a running count of the 'g'-names that occurred to her, and so on, are causally related in 'the right way' to her intention to *B7*. Of course, providing an account of 'the right way' has proved to be a challenging task. Paul Moser and I have addressed this issue in constructing an analysis of intentional action (Mele and Moser 1997); and I will not retrace our steps here. However, one feature of our analysis is especially significant for the purposes of this article. It merits attention.

When an agent does something intentionally, on our view, the acquisition of a proximal intention does not merely trigger a response; the continued presence

[7] In this paragraph and the next two, I borrow from Mele 1997: 236–7.

[8] Here, to keep things stylistically simpler than they would otherwise be, I use 'this intention' as shorthand for the following: this intention, its acquisition or persistence, or the neural realizer of any of the preceding. Reference to neural realizers is suppressed in the ensuing discussion.

of the intention also *causally sustains* the response if the response is more than a momentary event (Mele and Moser 1997: 234–7; cf. Mele 1992: 130–1, 180–1, 192–4). If a golfer, during his backswing, were to abandon his intention to hit the ball with his nine iron, he would halt his swing and select another club. Gail's continued trying is sustained, I suggest, by a persisting intention to complete the task of bringing it about that she thinks of seven animal names starting with 'g'.

Did Gail *intentionally* bring it about that she thought of seven animal names starting with 'g'? (That is, did she intentionally *B7*?) What it is to do something intentionally is a matter of considerable debate. However, if Gail's *B7*-ing is an action and *B7*-ing is something that she tried at the time to do, then given 'the absence of deviant causal chains and other monkey business' (Mele 1997: 234), her *B7*-ing is an intentional action, provided that her control over the success of her attempt does not fall short of the control required for intentional action. Now, part of the debate over how intentional action is to be analyzed is a debate over the control required for intentional action. Although Christopher Peacocke asserts that it is 'undisputed' that an agent who makes a successful attempt 'to hit a croquet ball through a distant hoop' intentionally hits the ball through the hoop (1985: 69), Brian O'Shaughnessy maintains that a novice who similarly succeeds in hitting the bull's-eye on a dart board does not intentionally hit the bull's-eye (1980: ii. 325; see Harman 1986: 92). The disagreement is about control, reliability, or skill. How reliable a bull's-eye hitter does one need to be in order to hit a bull's-eye intentionally? Whatever the answer to that question may be, we may suppose that it was not difficult for Gail to *B7*, and, indeed, that if she had chosen to spend a little more time on the task, she would have thought of more animal names that start with 'g'. In fact, Gail's *B7*-ing was no harder for her than my cooking a normal dinner is for me. (For your information, the great majority of dinners I cook, I cook intentionally.) The claim that Gail's *B7*-ing is an intentional action looks very plausible.

How might one attempt to undermine the claim that Gail's *B7*-ing is an intentional action? One strategy is to argue for the assertion (*A1*) that, regarding each name that ends up on Gail's mental list of seven 'g'-names, her bringing it about that she thinks of that particular name is not an intentional action and then to try to build a bridge from *A1* to the claim that her *B7*-ing is not an intentional action. Of course, the bridge should not be as simple as the following: Gail's *B7*-ing is composed entirely of seven different bringings-about—her bringing it about that she thinks of 'giraffe,' her bringing it about that she thinks of 'goat,' and so on—and none of those seven bringings-about is an intentional action; so her *B7*-ing is not an intentional action. Lots of arguments of that form are clearly flawed. (Example: *P* is composed entirely of seven different pieces, and none of those pieces is a jigsaw puzzle; so *P* is not a jigsaw puzzle. In fact, *P* is a seven-piece jigsaw puzzle for young children.) Also, if I am right, Gail *B7*-s at least partly by doing such things as keeping her attention focused on the task of

B7-ing and silently repeating to herself the 'g'-names that have come to mind as her mental list grows. These are intentional actions.

How might one argue for *A1*?[9] According to the so-called 'Simple View' of the connection between intention and intentional action, it is a necessary truth that any agent who intentionally *A*-s has 'an intention to *A*' (Bratman 1987: 112). Suppose that the Simple View is true and that (to put it roughly: see n. 8) intentions are causes of actions that execute them. Does this compound supposition permit someone's bringing it about that she thinks of 'goat' to be an intentional action? Of course not. In acquiring the intention to bring it about that one thinks of 'goat,' one would have thought of 'goat'.[10] This thinking of 'goat' would not be the execution of an intention to bring it about that one thinks of 'goat'.

I have argued elsewhere, as have others, that the Simple View is false.[11] Someone who is persuaded by these arguments may try to support *A1* by arguing that bringing it about that one thinks of 'goat' also is not an intentional action on the model of any overt actions that are counterexamples to the Simple View. The literature on the topic offers alleged counterexamples of three types: unproblematically successful tryings to perform an overt *A*-ing that are not associated with intentions to *A*; intentional overt 'side-effect' actions; and overt intentional *A*-ings that are routine actional parts of larger overt intentional actions performed in the absence of intentions to *A*.

Elsewhere, I have argued that an agent who is not in a position to intend to *A* may nevertheless be in a position to intend to try to *A*, and if that intention smoothly issues in an *A*-ing, the agent may properly be said to have *A*-ed intentionally (Mele 1992: 132–5; see Bratman 1987: 121). I would offer an illustration, if it were not obvious that Gail cannot intentionally bring it about that she thinks of 'goat' on a model that features intending to try to bring it about that she thinks of 'goat'.[12] In acquiring an intention to try to bring it about that she thinks of 'goat,' Gail would have thought of 'goat'.

One view in the literature on double effect is that agents perform some intentional actions that they neither intend nor try to perform. Consider the following from Gilbert Harman: 'in firing his gun,' a sniper who is trying to kill a soldier 'knowingly alerts the enemy to his presence' (1997: 151). Harman claims that, although the sniper 'does not intend to alert the enemy,' he *intentionally* alerts them, 'thinking that the gain is worth the possible cost'

[9] For readers who wonder why I raise this question despite having made my point about the simple bridge, I observe that not all potential bridges from *A1* to the thesis at issue are so simple. Also, discussion of *A1* will prove instructive.

[10] As I understand intentions, 'goat' would enter into the representational content of the alleged intention. So e.g. a Spanish speaker who is trying to bring it about that he remembers the English word for 'la cabra' does not have an intention to think of 'goat.'

[11] See e.g. Mele 1992: ch. 8, and Bratman 1987: ch. 8. For defenses of the Simple View, see Adams 1986 and McCann 1986.

[12] Bratman's well-known 'video games' thought experiment is an illustration (1987: ch. 8).

(p. 151). Michael Bratman makes a similar claim about a runner who reluctantly wears down some heirloom shoes (1987: 123; see Ginet 1990: 75–6). Obviously, Harman's sniper does not try to alert the enemy, nor does Bratman's runner try to wear down his shoes. Rather, they foresee that they will do these things as a consequence of doing something that they intend to do and try to do. Perhaps Harman and Bratman are right about this. If the sniper's alerting the enemy and the runner's wearing down his shoes are intentional actions, then they are actions in a sense of 'action' that philosophers of action take seriously.

Of course, anticipated, intentional 'side-effect' actions do not provide a model for understanding Gail's bringing it about that she thinks of 'goat' as an intentional action. It certainly is not the case that she anticipated that, as a consequence of trying to bring it about that she thinks of seven animal names starting with 'g,' she would think of 'goat.' In anticipating that, she would have thought of 'goat.'

The third type of exception to the Simple View features routine overt actional parts of intended intentional overt actions (Mele 1997: 242–3). Perhaps each of the steps I take when walking from my car to my office is an intentional step, even if I do not have a separate intention for each step. The steps might be intentional actions in virtue of being routine actional parts of an intended intentional action—my walking to my office (Mele 1992: 183–4). Is Gail's bringing it about that she thinks of 'goat'—her *BG*-ing—an intentional action in virtue of being an actional part of her intended intentional action of bringing it about that she thinks of seven animal names starting with 'g' (her *B7*-ing)? That requires that her *B7*-ing is an intentional action, and that her *B7*-ing is an intentional action is the main claim at issue. Simply assuming that her *B7*-ing is an intentional action would be question-begging in the present context. The goal of the imaginary defender of *A1* is to show that Gail's *B7*-ing is not an intentional action. A disanalogy between the two cases also merits mention: bringing it about that she thinks of seven animal names starting with 'g' is not routine behavior for Gail.

The preceding discussion of *A1*—the claim that, regarding each name that ends up on Gail's mental list of seven 'g'-names, her bringing it about that she thinks of that particular name is not an intentional action—suggests that *A1* should be taken seriously. Against *A1*, it may be claimed that, for example, Gail's *BG*-ing is an intentional action in virtue of the following fact: not only was her *BG*-ing a nondeviant product of her intention to *B7*, but it was a product that was appropriate to contributing to the satisfaction of that intention. But this will not do. Consider Loretta. Her task is to draw seven marbles from a vat, one at a time, while blindfolded—to *D7*, for short. She does this on television in her capacity as an employee of the Nevada State Lottery. Each marble is numbered. The numbers on the marbles she draws on this occasion are, in order, 17, 09, 98, etc. Although Loretta intentionally *D7*-s, she does not intentionally draw the 17

marble from the vat, nor the 09 marble, and so on.[13] (If she had intentionally drawn the 17 marble or the 09 marble, she would have been cheating; and Loretta was not cheating.) Loretta's drawing the 17 marble was a nondeviant product of her intention to *D7* and it was appropriate to contributing to the satisfaction of that intention. Even so, it is false that she intentionally draws the 17 marble.

Someone concerned to argue that Gail's *BG*-ing is an intentional action may try to exploit a difference between the two cases. Gail's bringing it about that she thinks of 'goat' is appropriate to contributing to the satisfaction of her intention to *B7* partly in virtue of its being '*goat*' that she thinks of, but the appropriateness of Loretta's drawing the 17 marble for contributing to the satisfaction of her intention to *D7* has nothing to do with its being the 17 marble that she draws.

Is Gail's *BG*-ing an intentional action? To be sure, she intentionally keeps her attention focused on the task of *B7*-ing, and as a nondeviant consequence of that she brings it about that she thinks of 'goat'—an appropriate word. Suppose that it was likely that she would *BG*. Are these conditions sufficient for Gail's *intentionally* bringing it about that she thinks of 'goat'? Proponents of the Simple View would certainly claim that the answer is *no*, and the following analogy supports that claim.

Ike is eligible for a prize on a game show. He will win the prize if and only if seven marbles with an animal name starting with 'g' printed on them are released from a vat. There are nine marbles in the vat, each of which has a different word printed on it. Seven of the words are animal names starting with 'g,' including 'goat,' and two are not. A marble will be released each time Ike presses a button. Pressing the button causes the marbles to bounce around in the vat until one lands in a special slot, at which time that marble is released. Ike, who does not know what words are printed on the marbles, is allowed to press the button only seven times. He intentionally presses it seven times, and as a nondeviant consequence of that he brings it about that the 'goat' marble is released.[14] Obviously, it was likely that he would bring that about. Even so, it would definitely be a stretch to say that he intentionally brought it about that the 'goat' marble was released.

That this is a stretch will not be noticed by someone who mistakenly reads 'he intentionally brought it about that *x*' as 'he intentionally did something, *A*, and his *A*-ing brought it about that *x*.' That this is a misreading is obvious. Don intentionally illuminated the room, and his illuminating the room brought it about that the prowler was alerted to Don's presence. But Don did not intentionally bring it about that the prowler was alerted to his presence.

[13] A Davidsonian would say that nothing that Loretta does is intentional under the description 'draws the 17 marble'.

[14] Of course, it is a single intentional button pressing that releases the 'goat' marble, not Ike's intentionally pressing the button seven times. But, in Gail's case, it is not Gail's keeping her attention focused on the task for the entire three minutes that has the consequence that she brings it about that she thinks of 'goat'. 'Goat' was the second 'g'-name she thought of.

(A Davidsonian would say that nothing Don did was intentional under the description 'bringing it about that the prowler was alerted to his presence.')

Suppose that Gail's *BG*-ing is not an intentional action and that *A1* is true. Can *A1* be used in a powerful argument for the thesis that Gail's *B7*-ing is not an intentional action? I have already offered grounds for accepting the claim that Gail intentionally *B7*-s, and I have identified problems for some ways of arguing from *A1* to the falsity of this claim about Gail's *B7*-ing. Loretta's case poses another problem. Even though Loretta does not intentionally draw the 17 marble from the vat, nor the 09 marble, and so on, she intentionally *D7*-s. Here is an analogous claim about Gail: even if she does not intentionally bring it about that she thinks of 'giraffe,' nor 'goat,' and so on, she intentionally *B7*-s. The analogy between Loretta's and Gail's cases makes it very difficult to argue from *A1* to the conclusion that Gail does not intentionally *B7*.

Here are some claims about Gail that I have defended in this section:

1. Gail's thinking of 'goat' (for example) is not an action.
2. Gail's thinking of seven animal names starting with 'g' (her *7*-ing) is not an action.
3. Gail's trying to bring it about that she *7*-s is an intentional action.
4. Gail's bringing it about that she *7*-s (her *B7*-ing) is an intentional action.

One point I made is that, in light of 3 and 4, fans of mental action should not be disheartened by 1 and 2.

The discussion of 4 can be used to reinforce the argument offered thus far for 2. Gail successfully tried to *B7*, and that successful attempt, is, I suggested, her *B7*-ing. The attempt was sustained by her intention to *B7*, and the attempt required for its success that Gail have conscious thoughts of seven animal names starting with 'g.' By analogy, one's attempt to raise one's right arm requires for its success that one's right arm rise. The fact that one's right arm rose is conceptually entailed by the fact that one raised it, and the fact that Gail had conscious thoughts of seven animal names starting with 'g' is conceptually entailed by the fact that she *B7*-ed. Now, whereas my raising my right arm is a candidate for being a basic action (see n. 6), Gail's *B7*-ing is not. She *B7*-s by performing a collection of actions, including keeping her attention focused on the task of *B7*-ing, focusing for a time on the more specific task of bringing it about that she thinks of fish names starting with 'g,' doing the same for insects, silently keeping a running count of the 'g'-names that occurred to her, and silently repeating to herself the 'g'-names that came to mind as her mental list grew. Thus far in this paragraph, nothing has been said about an action of *7*-ing. Does that entail that I have omitted something? I think not. I have said enough for the reader to understand how Gail brought it about that she had conscious thoughts of seven animal names starting with 'g,' and 'Gail's thinking of seven animal names starting with "g"' is plausibly regarded as shorthand for 'Gail's having conscious

thoughts of seven animal names starting with "g," " an expression that picks out a nonactional analogue of my arm's rising in a scenario in which I intentionally raise my arm.

4. CONNECTING WITH A RECENT DISPUTE

In Mele 1997, as I mentioned, I claimed that someone 'performed an action describable as "thinking of . . . seven kinds of animal having a name starting with g" ' (p. 234). Here, I have argued that thinking of seven such names is not an action but that Gail's bringing it about that she thinks of seven such names is an action—indeed, an intentional action. Should this—and, more to the point, the supporting discussion thus far—interest anyone but me?

Galen Strawson writes: 'I have argued that there is very little action in mental life, especially in the case that most concerns me: cognition in the widest sense. No coming to entertain a content, and no comprehending entertaining of a content, in reasoning, thinking, judging, or anything else, is itself an action' (2003: 244). Andrei Buckareff complains that 'Strawson fails to come close to sketching a theory of action . . . that could aid in the task' of distinguishing mental actions from nonactional mental events (2005: 84), and he contends that 'a good deal more mental activity than Strawson admits is actional' (p. 83). Since Strawson reports that he uses ' "action" to mean intentional action' (p. 228), Buckareff's use of 'actional' here should be understood accordingly. The preceding sections bear directly on their disagreement.

Consider Strawson's claim that 'No coming to entertain a content, and no comprehending entertaining of a content . . . is itself an action' (2003: 244). If 'to entertain a content' *c*, as Strawson intends that expression to be understood, is just to have a conscious thought the content of which is *c*—for example, a conscious thought of the word 'goat' or of a proposition expressed by 'goats are hairy'—then it is natural to understand 'coming to entertain a content' in such a way that it itself is never an action. Indeed, the expression last quoted can be definitionally reserved to designate something that happens to one. But when we ask why a particular instance of this happens to someone, we may find that mental actions are an important part of the answer. Why did Gail come to have a conscious thought of 'goat'? Partly because she tried to bring it about that she thought of seven animalnames starting with 'g.' And, as I have observed, she tried to do this by, among other things, keeping her attention focused on the task, silently keeping a running count of the 'g'-names that occurred to her, and silently repeating to herself the 'g'-names that already came to mind each time she thinks of another one. All of the actions just named are intentional mental actions.

What about a 'comprehending entertaining of a content'? Mike, who knows no Spanish, sees 'la cabra' on a store window while visiting Barcelona. The word attracts his attention. He has a noncomprehending conscious thought of

'la cabra.' He asks his Spanish guide, Nancy, what it means, and she says 'goat.' Nancy has a comprehending conscious thought of 'la cabra.' Plainly, neither her having that thought nor her coming to have it is an action.

Recall that Strawson says that by 'action,' he means 'intentional action.' Even if coming to have—and having—a (comprehending) conscious thought of a content *c* is never an intentional action, that leaves plenty of room for related intentional mental actions. Gail's intentionally bringing it about that she thinks of seven animal names starting with 'g' (intentionally *B7*-ing) is a case in point. In doing that, she brings it about that she comes to have—and therefore has—various comprehending, conscious thoughts that are relevant to her task. That these things that she brings about are not actions does not stand in the way of her *B7*-ing being an intentional action.

Return to some of the mental actions mentioned in my introduction. In intentionally solving a chess problem in one's head, one brings it about that one comes to have—and therefore has—a comprehending, conscious thought of the (or a) solution; and the same is true of an agent's solving 'a complicated arithmetic problem' in his head. In the course of intentionally deliberating about whether to accept a job offer, one may bring it about that one comes to have a great many comprehending, contentful, conscious thoughts—thoughts about various pros and cons, for example, and about their importance to one.[15]

These instances of coming to have a comprehending, contentful, conscious thought are analogues of some nonactional parts or products of overt intentional actions. In intentionally writing seven animal names beginning with 'g' on a sheet of paper, I bring it about that seven such names are on the sheet. The names' coming to be on the sheet is analogous to Gail's coming to have comprehending, conscious thoughts of various animal names starting with 'g.' After my initial deliberation about whether to accept a job offer, I may intentionally write down all the pros and cons that I remember considering. Their coming to be on my written list is analogous to my coming to have comprehending, conscious thoughts of them during my deliberation. In intentionally putting a dish in the dishwasher, I bring it about that the dish comes to be in the dishwasher. Its coming to be there is analogous to the mental problem-solver's coming to have a conscious thought of the solution to his problem. These analogies are not perfect. For example, I intentionally write each and every animal name I write on the sheet of paper, and Gail apparently does not intentionally bring it about that she thinks of 'goat' or any of the particular 'g'-names that end up on her mental list. But the

[15] I also mentioned deciding to *A* in my introduction. I examine that phenomenon at some length in Mele 2003: ch. 9, and I will not discuss it here. Strawson writes: 'most deciding what to do is best seen as something that just happens' (2003: 244). If he means something that just happens to the person who is said to decide, as opposed to something that the person (intentionally) does, a distinction between the nonactional acquisition of an intention to *A* and the mental action of forming an intention to *A* is relevant (Mele 1992: 141, 231; 2003: ch. 9). If Strawson were to accept this distinction, he might claim instead that most intention acquisition is nonactional.

point is that, in both cases, actions need to be distinguished from nonactional parts or products of actions.

There are tighter analogies. Recall Ike, the game show contestant. By intentionally pressing a button seven times, he brings it about that the 'giraffe' marble, the 'goat' marble, and so on, are released from a vat. Ike's pressing the button seven times is an analogue of Gail's trying to bring it about that she thinks of seven animal names starting with 'g,' his bringing it about that 'the goat' marble falls from the vat is an analogue of her bringing it about that she thinks of 'goat,' and the 'goat' marble's falling from the vat is an analogue of Gail's coming to have a comprehending, conscious thought of 'goat.'

The only example of mental action that Buckareff discusses in any detail in his critique of Strawson 2003 is an instance of mental arithmetic (2005: 85). I will discuss a mental arithmetic example of my own. I mentioned that some instances of blinking are actions and some are not. The same may be true of mental addition, depending on how that notion is to be understood. Like many people, Olive memorized lots of sums in school. When someone asks her what $6 + 4$ equals, or $18 + 7$, $29 + 6$, or the like, she knows straight off. Either she remembers the answer or a rapid computation of which she is not conscious generates it. If mental addition is understood in such a way that Olive's remembering the answer in these cases counts as mental addition, then some instances of mental addition are not actions. The same is true if mental addition is understood in such a way that Olive's unconsciously and effortlessly computing the answer counts as mental addition and if Buckareff (2005: 85) is correct in holding that such instances of computing are not actions. But other instances of mental addition are actions. Just now, Olive's young son asked her what $887 + 145$ equals, and she computed the sum in her head. Her strategy was to round 887 up to 900, add 145 to that, and subtract 13 (which she did by subtracting 10 and then 3 from 1045). If she had been thinking aloud—voicing her most salient conscious thoughts during the process—she might have said '900, 1045, 1035, 1032.' She did not consciously compute $900 + 145$ (though someone with less practice might have), but she did consciously compute $1045 - 13$ (in two steps, though someone with more practice might have done it in one).

Olive's having a comprehending conscious thought of the answer to her son's question is not an action; but she consciously tries to compute the answer, and she succeeds. The strategy she uses is a reasonable one, she is a competent user of it, and her success does not depend on any deviant causal chains. There is nothing to stand in the way of the judgment that Olive's computing the answer is an intentional action.

Strawson writes:

> The central point is this: the role of genuine action in thought is at best indirect. It is entirely *prefatory*, it is essentially—merely—*catalytic*. . . . There may well be a distinct,

and distinctive, phenomenon of setting one's mind at the problem, and this phenomenon, I think, may well be a matter of action. It may involve rapidly and silently imaging key words or sentences to oneself, rehearsing inferential transitions, refreshing images of a scene, and these acts of *priming*, which may be regularly repeated once things are under way, are likely to be fully fledged actions. [Also] sometimes one has to shepherd or dragoon one's wandering mind back to the previous thought-content in order for the train of thought to be restarted or continued, and this too may be a matter of action. . . . No doubt there are other such preparatory, ground-setting, tuning, retuning, shepherding, active moves or intentional initiations. But action, in thinking, really goes no further than this. (2003: 231–2)

The claims Strawson makes here may help to explain why Buckareff appeals to mental arithmetic. Olive does more than the applicable things mentioned in this passage: for example 'setting [her] mind at the problem.' She also consciously computes the answer. So Strawson goes too far. He would do well to observe that Olive's coming to know the answer to her son's question is not an action. But Olive does bring it about that she knows the answer by computing the answer, and her computing the answer is an intentional action.

Return to Strawson's thesis that 'there is very little action in mental life' (2003: 21). Some readers may take my discussion of a few representative examples to suggest that there is a lot of action in mental life. But, of course, even if you are persuaded by what I have written, its effect on your attitude about whether there is 'very little' or 'a lot' of action in mental life will depend on what you had been thinking about mental action. If you thought (perhaps unreflectively) that someone's remembering *x*, someone's thinking of *x*, someone's believing that *p*, and the like, are actions, it may now seem to you that there is relatively little mental action. However, as I hope I have shown, even if none of these things are actions, there is a lot of room for mental action in their production.

5. OBJECTIONS AND REPLIES

I conclude by considering some objections. The first concerns such assertions of mine as the following four:

1. Perhaps Gail tried to bring it about that she *7*-ed and succeeded in that, but did not, strictly speaking, try to *7* and did not perform an action of *7*-ing.
2. Gail's thinking of an appropriate 'g'-word should not be confused with anything she does that helps to bring it about that she thinks of that word; for events should not be confused with their causes.

3. Gail's *7*-ing should not be confused with the combination of her *7*-ing and things she does that help to bring it about that she *7*-s.
4. 'Gail's thinking of seven animal names starting with "g"' is plausibly regarded as shorthand for 'Gail's having conscious thoughts of seven animal names starting with "g,"' an expression that picks out a nonactional analogue of my arm's rising in a scenario in which I intentionally raise my arm.

The objection features two claims. First, such sentences as 'Gail thought of seven animal names starting with "g"' are ambiguous. Second, one reading of them is the one I motivated in this article, but another permissible reading treats Gail's thinking of seven animal names starting with 'g' (her *7*-ing) as the combination of her becoming conscious of a collection of seven such words and things she does that help to bring it about that this happens. This combination is Gail's bringing it about that she *7*-s (her *B7*-ing); and that, I have argued, is an action. The objector is contending that on my permissible reading of the quoted sentence, Gail's *7*-ing is not an action, but on another permissible reading, it is an action: namely, Gail's *B7*-ing.

That is fine with me. If by (*G1*) 'Gail's thinking of seven animal names starting with "g"' some people mean what I mean by (*G2*) 'Gail's bringing it about that she thinks of seven animal names starting with "g,"' I have no objection to their claim that the event they have in mind is an action. After all, I agree that it is an action. But if the thrust of the present objection is that my discussion of Gail's case makes no progress, I object. My account of Gail's intentionally *B7*-ing (or what some people might refer to as Gail's *7*-ing) sheds light on the nature of intentional mental action and has the potential for contributing to the refinement of disputes about how far the range of mental actions extends. If that account helps us to understand expressions like *G1*, as some people use them, so much the better.

I announced that 'Remembering that *p*, as I understand the phenomenon, is never an action; and the same goes for remembering *x* (e.g., a face, an odor).' It may be objected that some events that clearly are actions are identical with rememberings. For example, Hal's intentionally reciting in his head a haiku that he knows—his *H*-ing, for short—is a mental action; and it may be claimed that his *H*-ing and his consciously remembering the haiku are the same thing.[16] Here one must be careful. I am happy to grant that Hal's *H*-ing conceptually entails his consciously remembering the haiku. But this does not commit me to granting that Hal's consciously remembering the haiku is an action. After all, your raising your arm—an action—conceptually entails your arm's rising; but your arm's

[16] In correspondence, Matthew Soteriou inquired about the silent recitation of a poem.

rising is not an action. Similarly, your running a non-stop quarter-mile entails your legs' being in motion for some time, but your legs' being in motion is not an action (nor are your leg motions actions).

Perhaps the objector is thinking that not only does Hal's *H*-ing entail his consciously remembering the haiku, but his consciously remembering the haiku also entails his *H*-ing. Does the second entailment hold? Suppose a neuroscientist probes a patient's brain and the patient reports that, when the neuroscientist did that, a haiku he once memorized ran through his head.[17] The patient says, 'You pulled that Haiku out of my memory. I didn't silently recite it to myself; you didn't cause a silent reciting by me. The familiar haiku popped up in my head like words quickly popping up one at a time on a TV screen.' This is coherent. Someone can consciously remember a haiku without *H*-ing, just as one's arm can rise without one's raising it. As I understand the place of Hal's consciously remembering the haiku in his *H*-ing, it is analogous to the place of my arm's rising in my raising it.[18]

It may be objected that, in ordinary speech, 'remember' is sometimes treated as an action verb. I agree. Sometimes, by 'I remembered what I had for dinner last night' or 'I remembered the first verse of a poem I once wrote,' people seem to mean something that I would put as follows: 'I made a successful attempt to bring it about that I remembered what I had for dinner last night' or 'I silently recited the first verse of a poem I once wrote.' I have no objection to people talking this way (even though it may mislead some philosophers), and I agree that successful attempts to bring such things about are actions, as are conscious, silent recitings of poems. For my purposes in this article, there is no need to argue about ordinary usage of 'remember.' (Similarly, I have no need to argue about ordinary usage of 'Try to fall asleep!' If some people mean by that what I mean by 'Try to bring it about that you fall asleep!', that is fine with me.) It is the phenomena that have been under investigation here that interest me.

[17] Wilder Penfield writes: 'When I have caused a conscious patient to move his hand by applying an electrode to the motor cortex of one hemisphere I have often asked him about it. Invariably his response was: "I didn't do that. You did." When I caused him to vocalize, he said, "I didn't make that sound. You pulled it out of me" ' (1975: 76).

[18] The objector may search for another example of the desired kind—an event that is clearly an action and is identical with a remembering. As Matthew Soteriou reminded me, we sometimes say such things as 'Yesterday I forgot to phone Ann to invite her to my party, but today I remembered to do it.' Suppose that, when Art says this, part of what he means to communicate is that he phoned Ann today. Is he also saying or presupposing that his remembering to phone Ann and his phoning her are the same event? I doubt it. Art's 'today I remembered to do it' is plausibly regarded as convenient shorthand for something along the following lines: 'Because I remembered that, although I intended to phone Ann, I didn't do it, I phoned her today.' Here, the speaker identifies an action—the phoning—and part of the action's explanation.

I close with a reply to a final objection. I claimed that Gail consciously remembered the word 'goat.' One may argue that, despite the specific memory-priming context of Gail's thinking of 'goat,' a word she learnt and has a history of using, she consciously thinks of it but does not consciously remember it.[19] Here is my reply: Go for it; produce the argument! Once the argument is in place, it can be assessed. If it is convincing, will it challenge or support the claim that Gail's thinking of 'goat' is not an action? We must wait and see.[20]

REFERENCES

Adams, Frederick (1986) 'Intention and Intentional Action: The Simple View', *Mind and Language,* 1: 281–301.

Adams, Frederick, and A. Mele (1992) 'The Intention/Volition Debate', *Canadian Journal of Philosophy,* 22: 323–38.

Audi, Robert (1999) 'Doxastic Voluntarism and the Ethics of Belief', *Facta Philosophica,* 1: 87–109.

Bratman, Michael (1987) *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press).

Buckareff, Andrei (2005) 'How (Not) To Think About Mental Action', *Philosophical Explorations,* 8: 83–9.

Davidson, Donald (1980) *Essays on Actions and Events* (Oxford: Clarendon Press).

Ginet, Carl (1990) *On Action* (Cambridge: Cambridge University Press).

Goldman, Alvin (1970) *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall).

Harman, Gilbert (1986) *Change in View* (Cambridge, MA: MIT Press).

——(1997) 'Practical Reasoning', in A. Mele (ed.), *Philosophy of Action* (Oxford: Oxford University Press).

McCann, Hugh (1986) 'Rationality and the Range of Intention', *Midwest Studies in Philosophy,* 10: 191–211.

Mele, Alfred (1992) *Springs of Action* (New York: Oxford University Press).

——(1995) 'Motivation: Essentially Motivation-Constituting Attitudes', *Philosophical Review,* 104: 387–423.

——(1997) 'Agency and Mental Action', *Philosophical Perspectives,* 11: 231–49.

——(2003) *Motivation and Agency* (New York: Oxford University Press).

——and P. Moser (1997) 'Intentional Action', in A. Mele (ed.), *The Philosophy of Action* (Oxford: Oxford University Press).

O'Shaughnessy, Brian (1980) *The Will*, ii (Cambridge: Cambridge University Press).

Peacocke, Christopher (1985) 'Intention and Akrasia', in B. Vermazen and M. Hintikka (eds.), *Essays on Davidson* (Oxford: Clarendon Press).

[19] If someone asks you your name, do you remember what it is? You might say 'No, I just know it,' or you might say 'Yes.' Not all people use the word 'remember' in exactly the same way. If, on the street (where no philosophers are about), someone were asked his name and replied—apparently sincerely—that he did not remember, the questioner should have grave worries about him. Amnesia is a likely diagnosis.

[20] I am grateful to Andrei Buckareff, Randy Clarke, Stephen Kearns, Matthew Soteriou, and two anonymous readers for comments on a draft of this article.

——(1999) *Being Known* (Oxford: Clarendon Press).
Penfield, Wilder (1975) *The Mystery of the Mind* (Princeton: Princeton University Press).
Proust, Joëlle (2001) 'A Plea for Mental Acts', *Synthese,* 129: 105–28.
Strawson, Galen (2003) 'Mental Ballistics or the Involuntariness of Spontaneity', *Proceedings of the Aristotelian Society,* 103: 227–56.
Wittgenstein, Ludwig (1953) *Philosophical Investigations*, tr. G. Anscombe (New York: Macmillan).

3

Judging and the Scope of Mental Agency

Fabian Dorsch

1. SOME DIVISIONS IN THE MIND

Try to conjure up a visual image of a sunny forest, or to suppose that Goethe once visited Stoos in the centre of Switzerland. Presumably, you will be able to comply immediately and easily, without having to do something else first, and without having to invest too much effort. But then, try to conjure up a visual perception of a sunny forest, or to form the judgement that Goethe once visited Stoos, just like that—that is, without resorting to additional actions, such as travelling outside of the city, consulting a biography of Goethe, or taking a perception- or judgement-inducing drug. Presumably, you will fail.

What these examples suggest is that there is a fundamental—though perhaps not necessarily strict—divide among the phenomena making up our mental lives. On the one side, there are our deliberate and straightforward mental actions and the mental episodes which they produce (and sustain). Conjuring up an image or supposing that something is the case should count—if anything should—as paradigm instances of deliberate mental agency, namely as instances of the activity of imagining something.[1] Very roughly, they are examples of agency because they rationally respond to and are guided and possibly justified by certain practical reasons (i.e. those provided to us by our desires or intentions to picture or suppose something); they constitute mental actions because they are aimed at the production of some mental phenomena (i.e. an image or a supposition), and because their performance does not involve bodily movements, but occurs exclusively within the mind; they are deliberate because they are done in full, attentive consciousness of the means, ends, and intended results involved; and they are straightforward—or 'light-fingered' (O'Shaughnessy 1980: 21–2)—in that they are not done by performing another action which deliberately exploits certain passive effects (such as those of emotional dispositions,

[1] I defend the view that imagining is indeed a special form of deliberate and straightforward mental agency in Dorsch (2009).

hypnosis, or judgement-inducing drugs). Indeed, our ability to imagine seems to reveal the extreme freedom which we may enjoy in our conscious mental lives. It is difficult to think of a domain of our agency with fewer restrictions or obstacles.

On the other side of the divide, there are the more passive mental phenomena, notably those of our mental episodes, the formation (and sustainment) of which is either not at all influenced by our mental agency, or only in a mediated—though usually still deliberate—way. Many episodes occur and disappear without any active involvement on our behalf. We are often overcome by feelings or sensations, surprised or annoyed by perceived smells or sounds, or find ourselves suddenly confronted with images, memories, or thoughts. Other mental episodes, however, are located in between deliberately formed images or suppositions and passively occurring feelings or perceptions, in that their occurrence or nature is somehow influenced by deliberate activity, but in a less straightforward or encompassing way than in the typical case of imagining. You may intentionally calm down yourself or improve your mood by actively imagining something relaxing or cheerful, such as walking through a quiet and sunny forest and attending to the pleasant and soothing sounds and smells of the environment. Or you may be able to decide when to judge an issue or to remember an appearance, and which issue to judge or whose appearance to remember. But despite the intended impact of mental or bodily agency on such episodes, they do not seem to allow for the straightforward control that we often, if not always, enjoy over what we imagine. In this respect, they are closer to episodes which occur completely passively.

My primary interest in this chapter is to get clearer about how, and where, best to draw the line between the straightforwardly active and the more passive aspects of our conscious minds. My main focus will thereby be on the question of whether judging can be as active as imagining: that is, whether we can form judgements in a deliberate and straightforward manner, or (as I will also say) voluntarily, or at will. The example above suggests that the answer should be negative, and that we cannot freely decide which particular proposition to endorse in a judgement concerning a certain issue.

Indeed, this opinion has been widely endorsed.[2] But it has not often been explicitly argued for. And if it has, the arguments have typically made use of the controversial idea that judgements (or beliefs) are intrinsically and normatively linked to truth. I am sympathetic to the view that there is no real need to take up, but good reason to try to avoid, any commitment to such a norm for

[2] See e.g. Williams (1973), Winters (1979), O'Shaughnessy (1980), Bennett (1990), Noordhof (2001), Engel (2002), Owens (2003), and Shah and Velleman (2005). Even thinkers, who are more sympathetic to the idea of voluntary formations of judgements or belief, often defend this possibility without accepting (or at least arguing for) the possibility of judgements or beliefs formed directly in response to the intention to do so (cf. Weatherson's (2007) discussion, and especially his contrast between voluntary and volitional agency).

judgements (cf. Papineau 1999; Dretske 2000). And I will in addition illustrate, though only briefly, that the normative approach to the involuntariness of judgements faces a serious and not always noted problem.

As an alternative, I will present an argument which refers to the ways in which we consciously experience judgements and instances of deliberate mental agency. What is important about this argument for the involuntariness of judgements is not so much its perhaps uncontroversial conclusion, but rather the particular way in which it tries to support it. For it highlights the impact which conscious experience has on—as well as what this form of awareness can tell us about—the formation of judgements and the scope of deliberate mental agency. The resulting account is thus meant to be a promising competitor for theories which account for judgements and mental agency primarily in normative or functionalist terms.

The issue of whether we can form judgements at will can, as already suggested, be framed in terms of the question of whether the deliberate formation of a judgement can sometimes be as active as the conjuring up of an image or the making of a supposition, or whether instead it belongs always to the group of less active mental phenomena, which includes not only the passive cases of perceiving, sensing, or feeling something, but also the somewhat active cases of changing one's mood or bringing about the occurrence of an episodic memory. What distinguishes in particular the two ways in which we can deliberately influence the occurrence or nature of some of our mental episodes is that one is (as I have called it) more straightforward, or less mediated, than the other. Before I discuss both the normative and the experiential approach to the involuntariness of our judgements, it is perhaps helpful to say a bit more about what judgements are, and what characterizes straightforward deliberate agency.

Judgements—including occurrent beliefs—are mental episodes. As such, they are part of the stream of consciousness and of the same general kind of mental state as feelings, thoughts, sensations, perceptions, and so on.[3] More specifically, judgements consist in the conscious taking of a propositional and conceptual content to be true, which means that they do not—like suppositions—merely represent things as being a certain way, but also make the claim that this is how things are.[4] This is one important respect in which judgements differ from the events which constitute their formation (e.g. acts of judging): only the former,

[3] What I label 'mental episodes', Wollheim calls 'mental states' (1984: 33–4). And I will use the latter expression to cover not only mental episodes or events, but also dispositional or standing states, such as beliefs, desires, and so on. Accordingly, my talk of 'states' is not intended here as involving any commitment to a certain ontological category—such as to their being states of affairs, rather than events, activities, or processes.

[4] This does not suffice to distinguish judgements from guesses. What is needed in addition is that it is constitutive of a guess, but not of a judgement, that it originates in the conscious practical or arbitrary choice of one or another from a certain range of propositions, none of which are sufficiently supported or forced upon us by the evidence available to us, but which may none the less be among the acceptable options of choice only because they all enjoy at least some evidential backing.

but not the latter, are instances of episodic and committal thought. Another is that, while events of forming a judgement are often, or perhaps even always, instantaneous, the formed judgements may remain in continuous existence for a considerable amount of time. Our judgement that the person opposite us is very attractive, for example, may stay in the background of our conscious mind during our evening-long conversation and constantly influence what we say to, and how we treat, her or him.[5] And the same is true of many of our judgemental endorsements of the propositions which we use as premisses in more complex instances of reasoning.

The contrast between straightforward and mediated deliberate agency, on the other hand, is best drawn—at least for the current purposes—in terms of the fact that, while the latter involves the instrumental reliance on certain epistemic or merely causal processes and their passive effects, the former does not. The relevant processes are thereby characterized by the fact that, once they are triggered by us and progress normally, they lead by themselves—that is, without further help or involvement of agency or other factors external to them—to the desired or intended outcome. And to instrumentally rely on such processes means here to employ them as means in relation to their sufficiency, once triggered and progressing normally, to bring about the desired or intended outcome. What this involves, more specifically, is that we take the respective processes to be instrumental in achieving our goal; that we take them to be so partly by recognizing their capacity to lead by themselves to the desired outcome; and that we try to act on our relevant desires or intentions by means of trying to trigger the processes in question. Accordingly, an instance of deliberate agency aiming at the production of a mental phenomenon is mediated—and not straightforward—just in case the agent attempts to achieve this aim by trying to employ an epistemic or merely causal process as a sufficient means for bringing about the desired mental phenomenon. Any other involvement of epistemic or merely causal factors is compatible with both kinds of deliberate mental agency.

Consequently, we may still rationally endorse a proposition in a guess—but presumably not in a judgement—even when the proposition lacks sufficient epistemic support (cf. Owens 2003).

The offered characterization of judgements is also meant to focus the current discussion exclusively on kinds of judgement which are truth-apt and to be formed and assessed in relation to epistemic reasons. Whether this includes normative judgements (e.g. about what one ought to do, how things ought to be, or what is good) or judgements linked to or identified with intentions (e.g. about what one will do) depends on whether the correct account of these judgements will understand them as truth-apt and epistemic, or rather as expressive or as practical. However, my hope is that our experience of them as being formed in response to reasons—whether these are epistemic or practical—is also incompatible with any experience of them as being formed in a straightforward manner; and that this provides us again with an argument for their involuntariness. See Pink (1996) for an excellent discussion of, as well as a slightly different argument for, the involuntariness of decisions, desires, and normative judgements which are formed—or at least meant to be formed—in response to practical reasons.

[5] Thanks to Kevin Mulligan for suggesting this example.

For example, when we deliberately attempt to recall the appearance of a certain person, we thereby set in motion certain epistemic mechanisms because we expect them to force a specific image onto us that is likely to be accurate. If we lacked this expectation, we would probably often not bother engaging our memory, but instead opt for our ability to imagine appearances. And something very similar happens when we decide to judge an issue on the basis of the evidence available to us: we thereby assume that the proposition, which the evidence will compel us to endorse due to some underlying epistemic processes, will probably be true.[6] Similarly, we can reasonably decide and try to cause a change in our mood by imagining a certain scenario only if we believe that performing the latter action is likely to be instrumental in bringing about the desired alteration of how we feel. Otherwise, there would be no reason for us to engage in the imaginative activity in response to our wish to alter our mood. And when we deliberately take a drug in order to cause in us certain hallucinations, we do this precisely because of our reasonable expectation that the respective causal mechanisms, thereby triggered by us, are very likely to lead to the occurrence of such hallucinations. In the absence of this expectation, we normally would not take the drug, or at least not with the aim of hallucinating.

Imagining something, in contrast, does not involve similar instrumental beliefs about and exploitations of epistemic or merely causal mechanisms. Although straightforward agency may allow for the influence of, or even conscious reliance on, epistemic factors, this kind of influence appears never to be mediated by an epistemic process which is—or, at least, which we instrumentally believe to be—sufficient on its own to lead to the occurrence of whichever mental phenomenon has been wished for. Our successful attempt at visualizing a sunny forest, say, may very well be informed by our knowledge of how sunny forests look like, or of what it would be like to see a sunny forest. But the influence of this kind of knowledge on the resulting visual image is not mediated by (an instrumental belief about) a rational process pertaining to epistemic rationality which, once actively triggered, is alone responsible for the occurrence of the image. In particular, we do not form the visual image in response to our

[6] What happens in these cases is perhaps that our sole reason to rely on epistemic processes is that we are interested in producing true representations, and that we take epistemic processes (and nothing else) to be truth-conducive. But instead, it might also be a fact that basic cases of reliance on epistemic mechanisms within mediated mental agency do not actually involve the instrumental understanding and employment of the respective epistemic mechanisms. When we, say, begin to act on our decision to recall the appearance of a friend, we perhaps might not have to possess or use the instrumental belief that a good way of achieving this goal is to actively trigger the respective underlying mnemonic mechanisms. If this should turn out to be true, the characterization of the difference between straightforward and mediated agency would have to be refined accordingly—for instance, by weakening the respective condition to the effect that only the possibility, but not the actuality, of instrumental employment of the processes is to be taken into account; or by supplementing it with the condition that, if epistemic factors are concerned, it is alone decisive whether these are in fact triggered by our active engagement and furthermore by themselves sufficient to bring about the desired outcome, once started and progressing normally.

current recognition of some epistemic reasons (as in the case of the formation of judgements on perceptual or inferential grounds). And the occurrence of the visual image is not the direct result of a mechanism meant to preserve a rational link to epistemic reasons recognized in the past (as in the case of episodic memories based on past perceptions).

Similarly, it may be true that, when we successfully visualize a sunny forest, our employment of our capacity to visualize sunny forests is causally responsible for the occurrence of the respective visual image. And this again may involve, or be grounded in, more fundamental causal chains. But in order to successfully act on our intention to visualize a sunny forest, we need not, and do not, conceive of the causal processes concerned as sufficient means. We may take the employment of our capacity to visualize sunny forests to be a necessary part of visualizing a sunny forest, and we may even understand what we are then doing as the action of visualizing a sunny forest by means of actively making use of the respective capacity. But we do not thereby think of the link between our employment of this capacity (or similar capacities, such as the capacity to visualize trees) and the occurrence of the visual image in both causal and instrumental terms (if we think of the link or its potential causal nature at all). In particular, we do not form the intention to use our capacity to visualize sunny forests in rational response to our intention to visualize a sunny forest and an instrumental belief that making active use of this capacity is likely to cause the occurrence of the desired visual image. In fact, we would not know what it would mean to act on the intention to use our capacity to visualize sunny forests, if not to act on the intention to visualize one or more sunny forests.[7]

2. THE NORMATIVE APPROACH

In the light of the preceding considerations, the main task in the formulation of an argument for the claim that judgements can be formed at will is the identification of a constraint on all possible judgements, which puts them beyond our deliberate and straightforward control: that is, which prevents them from being consciously formed by us in direct response to our desire or intention to form them, without any instrumental exploitation of passive processes. If the constraint would not apply to all possible judgements, it could not completely account for the fact (assuming that it is one) that none of them can be formed by us at will, including the unconstrained ones. Furthermore, the constraint has to concern the ways in

[7] Note also that the case of visualizing someone by means of visualizing a portrait of this person still counts as an instance of straightforward agency, given that visualizing the portrait is neither epistemic, nor merely causal in nature, but instead itself a straightforward action. As a result, not all instances of straightforward agency need to be basic in the sense of not being performed by doing something else.

which judgements can be actively formed: it has to limit these ways in such a manner as to rule out the possibility of a deliberate and straightforward formation of judgements.

The most prominent strategy has been to derive such a constraint from the assumed fact that judgements are normatively linked to, or aim at, truth in such a way that they are subject to the following truth norm:[8]

(TN) Judgements ought always to be true, and to be formed only if they are true.

This truth norm is usually introduced for very different purposes, such as to capture the essence of judgements (or beliefs), or to account for their representationality and their link to truth (cf. Papineau (1999) and Dretske (2000) for a detailed discussion). That it may also figure in an account of the involuntariness of judgements is often only noted as an aside—if it is noted at all. But the truth norm is none the less predestined for playing this particular role because, when formulated in terms of (TN), it purports to achieve two things: to govern all possible judgements; and to put a restriction on the ways in which we can deliberately form them.

But not just any understanding of the truth norm and its impact on judgements can help in explaining the involuntariness of judgements. In particular, it does not suffice to identify the constraint on judgements and their deliberate formation, as it arises out of their assumed subjection to (TN), with the demand that, when deliberately forming a judgement (in contrast to another kind of mental episode), we should act on the aim to form the respective mental episode only if it is true. According to this demand, it is better or more appropriate to pursue truth as one's goal when deliberately forming a judgement. But it is not necessary, given that the possibility of violating the demand is not ruled out. Although one would be somehow at fault or irrational when ignoring or not following the demand, whether one satisfies it has no influence on whether one counts as deliberately forming a judgement. As a consequence, the demand does not really limit the ways in which judgements may be deliberately formed, it puts a restriction solely on when such an intentional formation may count as proper. It is, accordingly, not strong enough to prevent the occurrence of judgements which are formed entirely at will.

Therefore, the constraint on judgements derived from (TN) has to be understood in stronger terms. The most natural way of strengthening the condition on how we can deliberately form judgements seems to be to modify it

[8] See, for instance, Williams (1973) and Shah and Velleman (2005) for defences of this strategy and the kind of normativity involved, and Winters (1979), Bennett (1990) and Engel (2002) for critical discussions. See also Burge's writings, Peacocke (1998), Wedgwood (2002) and Shah (2003) for endorsements or explications of the idea that beliefs are subject to a truth norm, and that conformity to this norm requires us to believe something only if it is true. It is not unlikely that they will be sympathetic with the normative approach to the involuntariness of judgements.

in such a way that its satisfaction becomes constitutive of the deliberate formation of a judgement, instead of merely rendering examples of it appropriate.[9] The result will be something like the following requirement:

(C) Deliberately forming a judgement requires acting on the aim to form it only if it is true.

Thus, if we do not have this goal in mind and do not actively and consciously try to achieve it, then we are not engaged in the deliberate formation of a judgement—although we still might be engaged in the intentional or active formation of a mental episode of another kind (e.g. a supposition), or experience the passive occurrence of a judgement.

Understanding the constraint in this way does indeed promise to establish its incompatibility with any potential straightforward voluntariness of judgements. It seems plausible to say that deliberately acting on the aim to form a mental episode only if it is true requires making use of truth-conducive means. And, it may be further argued, reliance on truth-conducive means renders the respective mental agency mediated in the sense specified above. The idea is that only the reliance on epistemic reasons is likely to result in the formation of a true mental state. For, the assumption goes, there do not appear to be truth-conducive means other than epistemic considerations. Hence, the requirement (C) comes down to the demand that the deliberate formation of judgements has to happen by means of passive—namely epistemic—processes: judgements have to be deliberately formed on the basis of epistemic reasons (if they are to be deliberately formed at all). It follows from this that we cannot deliberately form judgements in a straightforward manner.[10]

But the constraint (C) does not apply to all possible or even all actual judgements. Paradigm examples of judgements, which are successfully formed in deliberate response to the desire or intention to form them, but without the aim in mind to form them only if they are true, are manipulative or induced judgements. Manipulated judgements are based on evidence, the collection of which involves ignoring evidence of a certain kind, or unproportionally or exclusively seeking evidence of another kind. Here are some good examples:[11]

Consider people who aim deliberatively to mislead themselves. Suppose an elderly man realises that he is likely to be upset if he learns about the real probability of his developing

[9] Williams (1973) can plausibly be read as adopting this strategy (cf. also Winters 1979).

[10] This argument is very similar to one of the arguments for the same conclusion in Williams (1973), only transposed from the conceptual level to the level of constitution (cf. my discussion in the last section). Proust (Ch. 13 below) presents a slightly different argument, the central idea of which seems to be that the aim of truth does not allow for the freedom of choice essential to deliberate agency.

[11] See O'Shaughnessy (1980), Owens (2000), Shah (2003) and Shah and Velleman (2005) for further examples of manipulated judgements, or 'wishful thinking'. Wedgwood (2002) also mentions the possibility of acting on one's intention to cease or avoid having a certain belief.

cancer, and so arranges to avoid any evidence that might undermine his sanguine belief that this probability is low. Or suppose an adolescent youth learns that people with an inflated view of their own worth are generally happier and more successful, and so deliberately seeks out evidence which will make him think overly well of himself. Of course, there are familiar psychological difficulties about deliberately arranging to have false beliefs, but examples like this suggest they are not insuperable. (Papineau 1999: 24)

There are probably many other, and possibly more radical, ways in which we can manipulate our evidence, other than by being unduly selective. For instance, we may ignore the lack of quality of some pieces of evidence (e.g. by relying on untrustworthy sources), or may invent or misread some of them (e.g. by misinterpreting emotional feelings as evidence). By contrast, induced judgements are formed in much simpler ways: they are not evidentially based, but instead occur as the product of some causal process which is intentionally triggered by some action of the subject in question. Examples of induced judgements would be those which would occur as the effect of the intake of a suitable drug, or of the visit to a hypnotist. They have in common with manipulated judgements that, often, they are deliberately formed without the aim of truth in mind. And although they may, as a consequence, end up being epistemically inappropriate, this does not undermine their possibility.

One might wish to insist that cases of manipulated or induced judgements do not really constitute counter-examples to (C), either because the mental episodes involved are not really judgements, or because they are not really actively or deliberately formed, so that their formation does not have to meet the necessary condition on the deliberate formation of judgements established by the constraint.[12] But it seems entirely ad hoc to claim that the examples do not concern judgements, given that the mental episodes in question endorse a proposition as true and are phenomenologically indistinguishable from more typically formed judgements (cf. Winters 1979 and Engel 2002). And the view that manipulated or induced judgements are not formed in an active and deliberate manner appears equally implausible. It is true that the agency involved leads to the occurrence of the judgement only in a mediated way. But something very similar is true of many other cases which we are normally happy to classify as deliberate actions. If bringing about the occurrence of a specific judgement by intentionally taking a drug in the full knowledge and reasonable expectation that this intake is likely to lead to the desired occurrence of the judgement is indeed

As O'Shaughnessy and Papineau observe, the intentional manipulation of one's judgements may require a certain amount of self-deception.

[12] In personal conversation, Shah mentioned that he is inclined to the view that the formation of manipulated or induced judgement should not count as an instance of (deliberate) agency. Indeed, he and Velleman seem to have to adopt this line of response, since they acknowledge the possibility of these types of judgements, but also believe that judgements cannot be formed at will because of their special normative nature described by (TN) (cf. Shah 2003; Shah and Velleman 2005).

not taken to constitute an action, then bringing about the death of a person by pulling a trigger or bringing about the arrival of a letter by posting it should not count as deliberate actions either.

Of course, the occurrence of the judgement is itself not an action. But neither is the occurrence of the death of the person, nor the arrival of the letter (at least not regarding the person who has sent it). Instead, what is actively done by the subject in question is the intended and expected bringing about of the occurrence of these passive events. And the subject performs this complex action—which may reasonably be described as the forming of a judgement, the killing of a person, or the sending of a letter—by performing a much simpler action, namely the taking of the drug, the pulling of the trigger, or the posting of the letter. It might still be attempted to maintain that, in general, there are no complex, but only simple actions; and that the latter do not allow for individuation and description in terms of their intended and expected causal consequences. But then, the formation of the judgement, the killing of the person, and the sending of the letter would still be on a par, since they all would equally not count as actions. And this result would fatally clash with our ordinary treatment of events of shooting someone or sending a letter—and not only of events of pulling a trigger or posting a letter—as instances of agency.[13]

The only significant difference between the two kinds of cases is that the occurrence of the judgement, but presumably not the occurrence of the death or the arrival of the letter, presents itself phenomenally to the agent as passive.[14] However, this is not the result of the judgement perhaps being brought about non-intentionally, or less actively than the two external events, but instead due to the fact that the judgement is part of the conscious mind of the subject and thus accessible to him in a different way than the external events. If he were able to become aware of the latter in the same way, he probably would experience them as passive as well. Moreover, the fact that the judgement is part of the subject's own mind, and not of another's, seems irrelevant for whether knowingly and expectantly bringing about of a judgement by, say, the administering of a drug should count as an action. None the less, this difference in how we are

[13] Even proponents of the view that only tryings are actions often enough permit that action descriptions can apply to complex events consisting in tryings and their causal results, as long as there is a suitable or non-deviant causal link between the two (cf. Hornsby 1980: 122–3; also O'Shaughnessy 1980).

[14] Another difference—though probably cutting across the cases—is that we do not always have established action terms available to directly pick out the more complex actions. We call the action of deliberately bringing about the death of a person by doing something simply a 'killing'. But there is no such action term for the action of deliberately bringing about the sleep of a person (e.g. oneself) by administering a drug to her. And the expression 'forming a judgement', as used for the action of deliberately bringing about the occurrence of a judgement by taking a respective drug, is probably not commonly understood as an action term. Note, however, that the last term, as well as the related expression 'the formation of the judgement', is meant here to pick out the event of doing something in order to cause the occurrence of a certain judgement, and not merely the event of this judgement occurring.

aware of judgements and external events may still ultimately explain why we may have the intuition that murdering a person or sending a letter have more right to count as actions than forming a judgement in one of the mediated ways mentioned.

It is conceivable that the truth-related normativity of judgements may give rise to requirements other than (C). But it is doubtful that any of them can be both weak enough to apply to all possible judgements, and strong enough to be incompatible with the deliberate and straightforward formation of judgements. In addition, the strategy of taking some constraint like (C) to be responsible for the involuntariness of judgements faces other difficulties, some of which I want to briefly mention.

A first challenge is to provide independent support for the claim that judgements are normative in the sense described. Critics of this idea have pointed out that it suffices for a satisfactory account of judgements (or beliefs)—which can explain, for instance, what differentiates judgements from other mental episodes, or how they represent the world—to assume that they have certain (evolutionary evolved) functions, notably the function to be true. This picture treats truth as a value for judgements, but as a value among many, which may be outweighed or undermined by the other values and thus need not always bind judges. That is, the latter need not always, when deliberately forming a judgement, be under the obligation to form it only if it is true. Accordingly, if the formation of a judgement is subject to such a truth-related obligation, this cannot be due to the general, intrinsic nature of judgements, but has to derive from something else, such as the wider practical purposes which are linked to the occurrence of the particular judgements in question, and which may differ greatly from case to case (cf. Dretske 2000; Papineau 1999). Another difficulty for the normative approach is to show how the requirement (C) can actually be derived from the truth norm (TN)—and if this fails, how it might be established on other grounds. And a third challenge is to demonstrate that intentionally grounding judgements in epistemic reasons is indeed a—and, moreover, the only—truth-conducive means available to us. In fact, it has been be doubted that deliberate reliance on epistemic reasons can function as an instrumental means to truth—at least, if the latter is to be understood as one of our purposes among many others (cf. Owens 2003).

3. THE EXPERIENTIAL APPROACH

That the normative approach arguably fails in its attempt to establish a constraint on our deliberate formation of judgements, which prevents it from being straightforward in all possible cases, provides a good reason to look for an alternative account. But the search for such an account is also, and independently, motivated by the reasonable expectation and hope that theories which deny (TN)

on other grounds should be able to account for the involuntariness of judgements as well as their norm-orientated competitors. My aim is therefore to pave the way for an argument showing why we cannot form judgements at will, which refers to the phenomenal character of judgements instead of their normativity.

This argument can be summarized as follows. Its starting point is the idea that we consciously experience our judgements always as epistemically motivated, while we consciously experience the straightforward results of our deliberate mental agency always as practically motivated. But, the reasoning continues, experiencing a mental episode as practically motivated rules out experiencing it as epistemically motivated—at least, if the episode concerned has been formed in a deliberate and straightforward manner. For experiencing such an episode as practically motivated means in fact experiencing it as immediately responding to the practical motives in question. And the phenomenal aspect reflecting this immediacy is incompatible with another potential aspect of experience, namely that aspect which reflects epistemic motivation. Hence, our judgements cannot result in a straightforward manner from our deliberate mental agency—which means that we cannot form them at will.

It will become much clearer in due course, I hope, how precisely each of the premisses involved in this argument should be understood, and also how they may be defended. But the core idea of this argument is that we experience certain conscious mental phenomena—such as judgements or mental actions—as rationally motivated. This means, first of all, that these phenomena possess a phenomenal or experiential character: they present themselves in phenomenal consciousness, or are experienced by us, in a specific way; or, as I will also say, they are phenomenally marked or revealed as being a certain way.[15] The core idea implies furthermore that the phenomenal characters of the phenomena in question are of a particular kind: they involve a rational dimension or aspect which reflects their rational nature. More specifically, the conscious mental phenomena concerned are phenomenally marked as standing in a certain kind of rational relation: we experience them as motivated by—that is, as rationally based on and occurring (or having occurred) in response to—reasons.[16]

[15] My use of the term 'experience' is perhaps unusual in that it refers to phenomenal consciousness rather than sensory experience. But it is akin to the German *Erlebnis* or *erleben* (especially as used by phenomenologists, such as Husserl) and will much simplify the presentation of the experiential approach. Other attempts at the notoriously difficult task of describing phenomenal consciousness have characterized it in terms of how it is or feels like to undergo, or be in, the respective events or states. Besides, I will leave it open whether the phenomenal character of episodes can remain unnoticed, or whether phenomenal consciousness always involves or requires some form of attention. This is unproblematic because forming a judgement deliberately, or 'in full consciousness' (cf. Williams 1973), will include attending to the judgement and the agency involved (cf. Peacocke (1998) and O'Brien (2003) for a discussion of this kind of attention).

[16] As I understand motivation here, it is equivalent to actual responsiveness to reasons, in the sense that a mental episode or event is rationally motivated if it is initiated, guided, or otherwise rationally determined by certain reasons. By contrast, in many meta-ethical discussions, the notion

In what follows, I will simply assume that judgements, mental actions, and the mental episodes which are the straightforward results of the latter are phenomenally conscious, or part of the stream of consciousness, and thus possess an experiential character. I will have to leave the defence of this assumption for another occasion.[17] Here, I will merely try to soften related doubts by making clear that assuming the experiential form of awareness at issue is less demanding than might be thought.

First, enjoying this kind of awareness need not require any specific conceptual capacities, even if describing it in terms of experiencing an episode *as* being a certain way might be taken to suggest just this. Saying that we experience certain mental episodes as responding to reasons does not mean more than saying that their phenomenal character shows a specific aspect, that the phenomenal character of other episodes lacks this aspect, and that this phenomenal difference somehow reflects the corresponding difference in origin and determination. In a similar way, we experience red-perceptions as representing a different colour than green-perceptions, or certain feelings as more pleasant than others. And although it should usually be possible for us to conceptualize such phenomenal differences in introspective higher-order judgements, such a conceptualization does not necessarily already happen on the phenomenal level.

Second, the form of awareness in question is minimal in the sense that we can experience a mental episode or event as rationally motivated without being aware of, or otherwise able to identify, the respective reasons. For instance, when asked what the capital of Ecuador is, we may form and rely on the judgement that it is Quito—say, as a manifestation of some previously acquired belief—without being able to remember when or how we learnt this fact (e.g. whether from listening to a teacher, from reading a book, or from looking at a map; cf. Wedgwood 2002: 20). And it commonly happens to us that we perform an action, such as entering a certain room, and recognize it as been done deliberately by us, although we have forgotten why we did it. Moreover,

of 'motivation' is used in a more narrow and perhaps more technical sense, being limited to what I call 'practical motivation'. Furthermore, I am not concerned with 'motivation' in the sense of having a certain desire or intention which has not (yet) become motivationally effective. And I also take it that there are important and phenomenally salient differences between epistemic and practical motivation—if only due to important differences between the respective kinds of reasons or rationality. One such difference is, for instance, that while practical ends may often be achieved in many different ways, reaching epistemic ends (i.e. truth or epistemic appropriateness) seems to always require the reliance on evidence. And while our various practical ends interact with each other (e.g. by outweighing or supporting each other), the epistemic ends appear to be completely independent and isolated from them (cf. Owens 2003).

[17] The assumption has been doubted, in particular, with respect to judgements. But most of the related debate has concentrated on whether judgements possess a distinctive phenomenal character, or whether differences in conceptual contents are phenomenally salient (cf. the discussions in Siewert 1998; Carruthers 2000), neither of which I assume here (cf. below)—though I defend, together with Gianfranco Soldati, both claims in Dorsch and Soldati (2004).

even when we are aware of the motivationally effective reasons and their specific nature, this awareness need not be experiential.

Third, experiential awareness may be fallible in at least two respects. In both cases, the phenomenal character of the episode or event concerned fails to adequately reflect its nature. But the reasons for this are different. On the one hand, this failure may be due to the fact that the episode or event in question does not live up to how instances of the mental type, to which it belongs, phenomenally purport to be. We may, for example, experience a judgement as a judgement and, hence, as responding to epistemic reasons, although it has been purely causally induced (e.g. by a drug, or by an emotion). On the other hand, the failure may stem from the fact that we erroneously experience an episode or event, not as an instance of the mental type to which it belongs, but as an instance of another type. We may, for example, experience a judgement as being a supposition—that is, we may experientially mistake a judgement for a supposition—and thus fail to experience the judgement as epistemically motivated.[18]

Correctly speaking, my main claim should therefore rather be that, if we experience a judgement as a judgement, or an instance or product of straightforward and deliberate mental agency as such an instance or product, then we always experience it as purporting to be rationally motivated. In other words, it is essential to how we experience episodes as being judgements, or alternatively as being part of straightforward and deliberate mental agency, that they present themselves phenomenally as occurring in response to reasons. But out of simplicity, I will continue to say that we always experience judgements, deliberate mental actions, and their straightforward results as rationally motivated.

And fourth, the conscious phenomena in question need not possess distinctive phenomenal characters, in terms of which they can be individuated and differentiated from other phenomena. Judgements, for example, need not phenomenally differ from other mental episodes which may also be experienced as epistemically motivated (e.g. perhaps, episodic memories); and they need not phenomenally differ among themselves, even if they differ, say, in content, origin, or motivation. All that is claimed is that judgements are experienced as supported by epistemic reasons. And similar considerations apply to our experience of deliberate mental actions and the mental episodes which they produce.

[18] The many examples of mainly pathological dissociations between our agency and our awareness of it (e.g. those mentioned in Wegner (2004) and discussed in his book) will also fall into one or the other category. The latter possibility of error would probably require, however, that episodes could be recognized and identified as judgements by reference to features other than their being actually experience by us in a certain way—for instance, in terms of their role in the acquisition of relevant beliefs or the performance of certain actions. And this might very well mean again that it is not essential to judgements that our actual experience of them shows some specific and distinctive aspects (though it may still be essential to them that they are consciously experienced in some way or another). More on the fallibility of experience and its relevance for the issue of involuntary judgements can be found at the end of sect. 6.

4. OUR EXPERIENCE OF JUDGEMENTS

What I want to try to defend first is the idea that the experiential character of our judgements always possesses a certain epistemic dimension: we experience our judgements as occurring in response to epistemic reasons. My defence makes essential use of an argument for the further claim that we experience judgements as epistemically reasonable, that is, as sufficiently supported by epistemic reasons.[19] This round-about strategy is possible because the two aspects of the experiential character of judgements concerned correspond to two intimately connected aspects of the epistemic status of judgements. Being reasonable is, for a judgement, partly a matter of being motivated by reasons, given that only (or at least primarily) motivating reasons contribute to the rational standing of a judgement. If I judge that something is coloured on the basis of my unjustified belief that it is green, my judgement will not be justified either, even if there is a motivationally ineffective, but good reason for forming this judgement available to me (e.g. the object may indeed be green, and I may generally be in the position to recall one of my correct past perceptions of it or simply to look at it again). Hence, if judgements turn out to be marked in phenomenal consciousness as reasonable, it is to be expected that they will also be phenomenally marked as rationally motivated. If it therefore can be made plausible that we are, in some way or another, aware of judgements as reasonable, this should provide substantial support for the claim that we are, in the same way, aware of them as being motivated by reasons. In particular, experiencing a judgement as rationally motivated would seem to be part of experiencing it as reasonable.

My argument begins with the observation that we take our judgements to be epistemically reasonable, at least as long as we are not aware of defeaters or do not otherwise begin to doubt the epistemic standing of the judgements in question. If we would not take our judgements to enjoy such reasonableness, we probably would not rely on them as a provider of reasons for belief or action, in the sense that we would not let them rationally contribute to our acquisition of the corresponding non-occurrent beliefs or, by means of further theoretical or practical deliberation, to our acquisition or revision of other judgements, beliefs, or intentions. Instead, we would be inclined to revise them or give them up, or indeed would have refrained from forming them in the first place. That we—at least initially—take our judgements to be reasonable becomes also apparent in cases in which we come to doubt the epistemic reasonableness of one of our already existing judgements—say, because we begin to question the

[19] In fact, these two phenomenal aspects seem to be part of an even richer epistemic dimension of the experiential character of judgements, consisting in our phenomenal awareness of them as *providing reasons* for belief or action.

quality of the supporting evidence, recognize some fault in the cognitive processes originally involved, or simply encounter an opposing view. The occurrence of such a doubt presupposes that we are already aware of an initial claim to reasonableness, which then becomes the subject of the doubt. In particular, doubting a judgement on the grounds, say, that the perceptual conditions are inadequate requires being aware of the fact that the judgement in question has enjoyed rational support by a perception had under those inappropriate conditions.

The observation that we take our judgements to be reasonable and, as part of this, to be rationally motivated fits well with two other observations, namely that we take our judgements to have occurred passively, and that we take them to amount to knowledge (or at least to purport to do so) and treat them accordingly—for instance, when we rely on them in the acquisition of beliefs or the formation of intentions—even if they do not constitute knowledge (cf. Williamson 2000; Wedgwood 2002; Hornsby 2005). It seems that we are aware of our judgements as passive precisely because of—and perhaps even by—being aware of them as based on epistemic reasons, that is, as determined by passive epistemic processes. And assuming that knowledge requires both truth and epistemic appropriateness, taking our judgements to be instances of knowledge appears to involve taking them to be both true and reasonable. Indeed, if it is furthermore accepted that the fact that judgements endorse a proposition as true, and thus make a claim about how things are, is phenomenally salient and distinguishes them experientially from, say, suppositions (cf. Dorsch 2005), it seems very plausible that the other aspect of the epistemic status of judgements—that is, their reasonableness—should also be perspicuous in this way. The idea is that, by presenting themselves in phenomenal consciousness as instances of knowledge (independently of whether they in fact amount to knowledge), judgements make two salient and interrelated claims to rationality: that they represent adequately how things are; and that they are thereby sufficiently rationally supported.

But a sceptic concerning the experiential awareness of the prima facie reasonableness of judgements is probably also a sceptic concerning the experiential awareness of their claim to truth and knowledge. Therefore, I would like to put forward another line of reasoning, according to which our primary awareness of the reasonableness of our judgements should be best understood as a form of experiential awareness, given that all plausible alternatives appear to be untenable. There seem to be two plausible competitors to this view: the inference model and the prompting model.[20] Both these models have in common that they take the awareness at issue to be the higher-order judgement that the

[20] Other candidates seem to be even less attractive (cf. O'Brien 2003), in particular the idea that the awareness in question is a matter of some internal form of perception or observation (cf. Shoemaker 1994; Burge 1996; Martin 1997).

respective lower-order judgement is epistemically reasonable. But they differ in their account of how we come to form that judgement.

The inference model maintains that the higher-order judgement under discussion is the result of a complex cognitive process. More specifically, it states that we infer the prima facie reasonableness of our judgements. For instance, we may believe that our judgements are generally reasonable as long as there are no relevant defeaters or doubts, and we may introspectively recognize that the mental episode in question is a judgement and that we have not been aware of any relevant defeaters previous to our doubts. Or, alternatively, we may remember how we have formed a judgement on the basis of certain pieces of evidence, and we may recognize that we have taken this formation to be epistemically appropriate, or at least have remained unaware of any inappropriateness, at the time of its occurrence. In both cases, we can then conclude that the judgement concerned has some claim to reasonableness.

However, that our primary awareness of the reasonableness of judgements is often not the result of such inferences is illustrated by cases in which we are ignorant about the general reasonableness of judgements, or about the particular epistemic origin of the judgement at issue. The view that our judgements are generally reasonable, as long as there are no defeating factors or circumstances present, seems to be sufficiently complex and non-obvious for many subjects (such as children) to lack it—in particular, since it requires a substantial amount of theorizing (assuming that it is not based on how we consciously experience judgements) and the possession of certain more technical concepts (such that of a defeater). But this does not seem to prevent those subjects from taking their individual judgements to be reasonable and to rely on them as providers of reasons for belief or action. Similarly, as already illustrated above, we may not remember what has ultimately provided support for our judgements and the beliefs which they may manifest. But we may still take them to be reasonable and trust them in our reasoning. And finally, our awareness of the reasonableness of our judgements often occurs too immediately to involve, or be preceded by, inferences of the kind described—for instance, when we enjoy such an awareness as part of coming to doubt the epistemic standing of a given judgement in direct reaction to, say, hearing a contradicting opinion or realizing that circumstances have been rather non-standard.

The prompting model, in contrast, claims that the higher-order judgement at issue (or a corresponding intuitive seeming[21]) occurs spontaneously, once we begin to wonder whether the lower-order judgement is reasonable. According

[21] The prompting model invites characterization in terms of intuitions. Depending on one's understanding of them, either the higher-order judgement itself, or some spontaneously occurring rational seeming, on which the higher-order judgement is directly based, may be said to be intuitive (cf. the essays in Bodrozic 2004)). In the latter case, my arguments against the prompting model will concern the spontaneously occurring seemings, rather than the higher-order judgements grounded in them.

to this view, the higher-order judgement is neither based on some inference or observation, nor simply the manifestation of a prior belief. Moreover, it is not based on the conscious experience of the reasonableness of the lower-order judgement. Otherwise, the experience would enjoy primacy over the higher-order judgement, and we would have the experience, rather than the prompting, model. Instead, the higher-order judgement is automatically prompted by our wondering about the epistemic status of the lower-order judgement in virtue of some reliable internal mechanism.

This internal mechanism cannot plausibly be due to some constitutive link between the lower- and the higher-order state. It does not seem to be true, for instance, that—assuming that we are rational and possess the required concepts—the presence of the lower-order judgement entails (and is perhaps entailed by) the possession of the higher-order belief that it is epistemically reasonable, or at least the willingness to form the corresponding higher-order judgement when considering the issue. The two mental phenomena in question seem to be of such kinds as to be much more distinct than that. We can be rational, have a well-functioning mind, and possess the concept of reasonableness (or even prima facie reasonableness), but, when asking which epistemic standing one of our judgements enjoys, still fail to apply the concept to the judgement. In particular, no aspect of this concept, or of our possession of it, tells us that it correctly applies to at least certain judgements. Recognizing that they enjoy such reasonableness amounts to a more substantial piece of knowledge.[22]

Hence, the lower-order judgement, together with our consideration of its epistemic status, is perhaps better taken to reliably give rise to the higher-order judgement via some contingent and merely causal or informational relation.[23] But apart from the general difficulties linked to causal reliabilism or informationalist semantics, this view faces the challenge to satisfactorily motivate the postulation of the respective internal mechanism. This mechanism would seem very odd and difficult to explain if such higher-order judgements spontaneously occurred

[22] The assumption of a constitutive link is perhaps plausible with respect to higher-order ascriptions of propositional contents or attitudes (cf. Shoemaker 1994; Burge 1996; Wright 1998). But judgements about the reasonableness of other judgements are clearly of neither kind. And see Peacocke (1998), Martin (1998), and O'Brien (2003) for more general objections to the constitutive account and its central claim that higher- and lower-order states are not distinct entities. One particular worry is, for instance, that the postulated constitutive link between the lower- and higher-order states does not seem to provide the resources to explain how the latter can be epistemically grounded on, and constitute genuine instances of knowledge of, the former.

[23] Peacocke's account of self-knowledge seems to open up a third possibility: to take the link to be rational, but non-constitutive. However, the considerations presented above against the applicability of constitutivist accounts also rule out the applicability of Peacocke's view, given that the latter, too, implies that the occurrence of the conscious lower-order states, together with our conceptual capacities and a rational and well-functioning mind, ensures that we are willing to form the higher-order judgements in the relevant circumstances (cf. Peacocke 1996). Besides, it is also interesting to note that Peacocke's account goes beyond the prompting model, and has some affinities to the experiential model, in that it assigns to consciousness an essential function in the epistemology of self-knowledge (cf. n. 28 below).

solely in response to wondering about the epistemic standing of lower-order judgements, and not in response to wondering about some other feature or some other mental episode. But to widen the scope of the prompting model to other kinds of higher-order judgements seems to be plausible only in the context of endorsing an account of introspection, or self-knowledge, in terms of contingently but reliably prompted higher-order judgements. And such an account seems to be highly implausible, especially if applied to the kind of awareness under discussion.[24] One specific problem is that the reasonableness of judgements seems to be among their features to which we can have direct introspective access only if they are indeed marked in phenomenal consciousness. The epistemic standing of judgements is at least in most cases a matter of their rational relations to reason-providing states or facts extrinsic to them. Therefore, if it is not reflected by an introspectible aspect of the experiential character of judgements, it can be recognized only by means of a cognitive process which encompasses more than the mere introspection of the judgement and its intrinsic features—a cognitive process which, for instance, combines introspection with inference and perhaps memory, as described above during the discussion of the inferential model.[25]

However, if our primary awareness of the reasonableness of judgements is in at least many cases neither based on inferences, nor a matter of causally prompted higher-order judgements, then it should be taken to be experiential. No other plausible alternative suggests itself. Hence, taking our judgements to be epistemically appropriate should be best understood as experiencing them as enjoying the support of epistemic reasons—which again involves experiencing the judgements as being motivated by such reasons. Our higher-order judgements about the reasonableness of our lower-judgements may then be the result of introspecting this epistemic aspect of the phenomenal character of the latter.

As I have already mentioned, that judgements are always marked in phenomenal consciousness as occurring in response to reasons is compatible with the possibility that they are actually not so motivated, and that correspondingly our experiential awareness has failed us. We presumably react to such cases of error by taking ourselves to have forgotten about the specific rational origins

[24] Among the more general objections to this account of introspection—which is endorsed, for instance, by Armstrong (1993) and Lycan (1996)—are: that it has to assume some form of causal reliabilism (cf. O'Brien 2003); that it cannot capture the transparency of mental content (cf. Dretske 1999); and that it does not link the lower- and higher-order states intimately enough to be able to account for the immediate rational impact of the latter on our revision of the former, for our related epistemic responsibility, and for the impossibility of brute error (i.e. error not due to the irrationality or malfunctioning of the subject) in the acquisition of self-knowledge (cf. Burge 1996; also Shoemaker 1994 and Siewert 1998).

[25] The same problem need not arise with respect to the introspection of externally determined contents, given that the contents of the higher-order states may embed the contents of the lower-order states and thus can inform us about them without having to tell us something about their extrinsic relations (cf. Burge 1996; Peacocke 1996, 1998).

of the judgements in question, or by coming to identify or construct new ones (e.g. by interpreting the mental causes of the judgements—say, certain desires or emotions—as their grounds). But our phenomenal awareness of judgements as rationally motivated is also compatible with the possibility of self-justifying judgements (if they are indeed a possibility). Such judgements provide epistemic support for themselves in virtue of some feature which they possess (e.g. their necessity, infallibility, certainty, etc.). And when we experience them as rationally motivated, we are aware of this rational relation in which they stand to themselves (e.g. by experiencing them as certain or self-evident). Nothing in what has been said so far suggests or even requires that we experience judgements as motivated by epistemic reasons distinct from themselves.

Much more problematic would be if some class of our judgements would allow for epistemic appropriateness despite not permitting any support by epistemic reasons, whether provided by the judgements themselves or by other states or facts. However, none of our judgements seem to be of such a kind.[26] From an epistemic point of view, such states would be much more similar to perceptions than to normal judgements. Maybe intuitions, or intellectual seemings, may be of this type. Just like perceptions, they can perhaps be reliable or otherwise epistemically appropriate, without standing in rational relations supporting them. And just like perceptions, they are perhaps also immune to any rational influence of reasons. But judgements seem to be very different. Their epistemic appropriateness appears to be partly a matter of how well they cohere with our already existing beliefs (as well as other judgements). And this seems to mean, among other things, that the latter may provide us with (access to) reasons for, or against, the formation or revision of judgements. Moreover, our judgements appear to be sensitive to such reasons and react accordingly—say, by disappearing when they are in too great a tension with what else we believe.

5. OUR EXPERIENCE OF MENTAL AGENCY

What is left to be shown is how their feature of being experienced as epistemically motivated prevents judgements from being formed at will. In the remaining sections, I will argue for this incompatibility in three steps. First, I will try to make plausible that deliberate and straightforward mental agency (if successful) results in mental episodes which are always experienced by us as actively formed. Second, I will argue that this actually means that the respective episodes present themselves as occurring in immediate response to practical reasons. And

[26] I take it that even perceptual judgements are not of this type, given that they are normally rationally supported by perceptions. See e.g. Martin (1993) and Cruz and Pollock (1999) for defences of this view.

third, I will show that no mental episode can be experienced by us both as being epistemically motivated and as being immediately practically motivated. From this incompatibility between the two ways in which we may consciously experience mental episodes, it follows that our conscious judgements cannot result in a straightforward manner from our deliberate mental agency: for us, there cannot be any judgements formed at will.

I take it that our instances of deliberate mental agency (and presumably of deliberate agency in general) normally involve at least three elements (cf. e.g. Pink 1996). First of all, there are certain practical reasons which are potential motives for action, and which we are put in contact with by some of our mental states—say, intentions, desires, or other states with the capacity to move us to act. Then, there are the mental actions themselves which occur when we begin to act on some of the provided reasons. Examples are the straightforward acts of conjuring up an image or of making an explicit assumption. The mental actions may thereby be partly or wholly successful in bringing about the respective mental phenomena; or they may amount to something like mere attempts or tryings. And finally, there is the motivational link between the two other elements. The mental actions come into being once our practical reasons actually begin to move us. And these reasons continue to guide us throughout our performance of the resulting actions (cf. O'Brien 2003). According to this picture, practical motivation is—just like epistemic motivation—a rational (and presumably causal) relation; and it obtains precisely as long as the practical reasons stay effective in initiating and guiding the mental actions concerned.

Often, however, our mental actions involve, or at least give rise to, a fourth element: they bring about certain desired or intended mental phenomena as their results. Trying to conjure up an image of a sunny forest may actually produce such an image; while attempting to improve one's mood by conjuring up such an image may result in one's becoming happier. Furthermore, some of these results may be due to deliberate and straightforward mental agency. The representational episodes produced by, or as part of, acts of successful visualizing or supposing—such as the image of the sunny forest—are good examples.

Now, we typically can tell whether one of our mental episodes has been the result of our mental agency—at least, if it has been produced in a deliberate and straightforward way. When you pictured to yourself the sunny forest, or supposed that Goethe went to Stoos, you were presumably aware of the fact that you actively formed the respective representational episodes. And your awareness of them would presumably differ in this respect from the awareness you would have when perceiving a sunny forest, or judging that Goethe visited Stoos, or being confronted with the spontaneous and unbidden occurrence of an image or thought with a corresponding content. The main issue with which I will be concerned in the remainder of this section is how we can come to acquire this kind of awareness. And the plausible options seem to be the same as in the case

of our awareness of the epistemic reasonableness of judgements: the inference, the prompting, and the experience model.[27]

As above, the inference model maintains that we inferentially arrive at our knowledge of the active origin of the respective mental episodes on the basis of introspection, and perhaps also memory or other forms of knowledge. A person successfully visualizing a sunny forest may, for instance, be introspectively aware of her intention—or of her attempt to act on her intention—to picture such a scene, as well as of the occurrence of the resulting image. Moreover, she may notice that these phenomena are temporally ordered, and that there is a match between the content of the intention or attempt at action and the nature of the subsequent visual episode. And she may possess general knowledge of the fact that such a combination of agreement and temporal order, which furthermore involves an intention or attempt to do something, is usually not accidental, but rather the consequence of the rational forces involved in practical motivation. Hence, the person may be able to draw the conclusion that the image of the forest has occurred, not spontaneously, but as the result of her own deliberate mental agency initiated by her intention.

However, the demands put by the inference model on the knowing subject are again too high. In order to come to know that some, but not others, of our mental episodes have been actively formed, we do not seem to have to possess knowledge of the non-accidental character and origin of the temporal order and match of the mental phenomena involved in mental agency. Nor do we seem to have to possess some of the concepts needed to entertain such knowledge or draw the inferences required (e.g. the concept of the kind of match described). In addition, our acquisition of the knowledge about the origin of our mental episodes appears, from a subjective point of view, to be more immediate than described by the inference model. It may be true that we infer the active or passive origin of a given episode in very special circumstances (e.g. when we are unsure about whether our primary way of acquiring this knowledge is working properly). But it seems that we typically do not need to engage in such elaborate reasoning in order to tell whether an episode is due to our own mental agency (cf. Peacocke 1998; O'Brien 2003).

According to the prompting model, the higher-order judgements about the active or passive origin of our mental episodes are not based on observation, inference, or experience, but instead reliably prompted by simply paying attention to the issue, or asking oneself the question, of how a certain present mental episode has been formed. Their occurrence is thus the product of some underlying causal or informational mechanism, which is set in motion by consciously addressing

[27] An observational model can again be ruled out straight away (cf. O'Brien 2003). Although it has been argued that our primary knowledge of our own bodily actions is mediated by proprioception (cf. Dokic 2003), this idea obviously cannot be applied to mental agency. For the same reason, outer perception could not play a role.

the topic of the origin of a given mental episode. And they reliably track the presence, or absence, of the special link obtaining between successful mental actions and the mental episodes which they have produced in a straightforward manner.

But this application of the prompting model is not very appealing, and mainly for the same reason as above, namely its difficulty in motivating the acceptance of the postulated internal mechanism. Again, it seems to make sense to speak of reliably prompted higher-order judgements only if they are taken to be introspective, or instances of self-knowledge. This idea is maybe more plausible this time, given that the straightforward results of our successful mental actions are perhaps constitutive parts of these actions (cf. Audi 1993), and introspecting the results and their active nature may therefore happen as part of introspecting the respective mental actions and their active nature. But it still seems valid that theories of introspection in line with the prompting model are much more plausible if they take the link between the lower- and the higher-order states to be constitutive rather than causal or informational (cf. n. 24 above). However, the possibility of a constitutive account does not arise, since we can satisfy all the relevant conditions concerning rationality, possession of concepts, and so on, without being inclined to judge a given lower-order mental episode to be actively or passively formed when asking ourselves the respective question.[28]

The experiential model seems, again, to be the best remaining alternative. It claims that we become aware of the active or passive origin of our mental episodes simply by having and experiencing them. In particular, we experience the mental episodes resulting straightforwardly from our deliberate mental agency as actively formed, while we presumably experience most or all other mental episodes as having occurred in a passive manner. This is one reason why we experience a

[28] See O'Brien (2003) for more general criticism of the constitutivist approach to our self-knowledge of our conscious actions. Her own account of such self-knowledge is formulated along the lines of Peacocke's account of our self-knowledge of conscious states. Accordingly, it assumes a rational, but non-constitutive link between our actions and our self-knowledge of them, as well as an essential role for the way in which we are conscious of our own actions, namely by means of a '[conscious] sense of guiding our action, . . . a sense of control' (ibid. 378). The latter aspect of her theory seems to be very close to my idea that we experience our mental actions as motivated and guided by practical reasons and renders her view more akin to the experiential than to the prompting model.

Peacocke's own view on our self-knowledge of (bodily) agency also assigns an essential epistemic role to our non-observational conscious experience (or 'awareness from the inside') in the formation of the higher-order judgement or belief that we are, or have been, successfully trying to do something, at least in the case of basic or straightforward agency. Moreover, he takes the respective experience to be an awareness of successfully trying (cf. Peacocke 2003: 103 and 105; and Ch. 10 below). Hence, the awareness may very well extend (as proposed by the experiential model) to that mental episode the occurrence of which renders the respective attempt at mental action successful. It is, however, unclear whether what he has in mind here is the way we phenomenally experience actions—i.e. a phenomenal property of the episodes of acting themselves; or instead independent and non-judgemental conscious states representing our mental actions (cf. Ch. 10).

deliberately formed image of a sunny forest differently from a perception or a spontaneous or remembered image of such a scene.

This picture fits very well with (but does not entail) the view that our primary awareness of the active character of our deliberate mental actions is experiential, too. And the truth of this further view would suggest (but, again, not imply) that the straightforward results of our deliberate mental actions, given that they are experienced in the same way as the mental actions itself, are constitutive parts of the latter. The fact that we experience certain mental phenomena as active may thus perhaps serve as a guide to agency: if we experience something as active, then it normally is an instance of agency. However, the opposite does not seem to be true: we do not appear to experience all instances or parts of action as active. Most, if not all, examples of non-deliberate agency seem to lack the kind of attentive conscious awareness of activity characteristic of deliberate agency (cf. O'Shaughnessy 1980; Pink 1996). And when we intentionally improve our mood by imagining something cheerful, we do not seem to experience the resulting change in mood as actively produced, but only the images and thoughts involved in bringing about that change. Given that the deliberate improvement of one's mood is none the less an instance of mediated agency, this suggests that our experience of passivity does not always reveal all aspects of the origin of the respective mental episodes. It discloses the direct passive origination in, say, some epistemic or merely causal processes. But it does not also reveal the prior deliberate mental activity which has started these processes. Therefore, this ultimately suggests that our experience of activity is solely or primarily a guide to deliberate and straightforward agency.

6. THE INCOMPATIBILITY OF THE TWO KINDS OF EXPERIENCE

This leads directly to the question of what it means to experience a mental phenomenon as active. My answer to this question is that the respective experience reveals at least two aspects of the mental agency concerned: that it is practically motivated; and that it is so motivated in an immediate manner. This may explain, among other things, why our experience of activity may very well be a guide to straightforward agency, assuming that, normally, the phenomenal character of our conscious mental states and events adequately reflects their nature.

It appears very natural to say that our mental phenomena, which are marked in phenomenal consciousness as active, present themselves thereby as practically motivated. Experiencing some action or episode as practically motivated means experiencing it as rationally responding to certain practical reasons. And our deliberate mental actions (including our not entirely unsuccessful attempts at them) are indeed sensitive to reasons in this way: they are initiated and guided

throughout their performance by practical reasons provided to us by our desires, intentions, or similar states. In fact, if our experience of our deliberate mental actions did not reflect this sensitivity to reasons, it would not make much sense to call it an experience of activity at all: practical motivation seems to be at the heart of agency. Not surprisingly, when theorists talk about how actions present themselves to us in phenomenal consciousness, they often resort to characterizations very similar to mine.[29] And of course, our experiential awareness of practical motivation is—just like our experience of epistemic motivation—possibly non-conceptual, minimal, fallible, and non-distinctive in the senses specified above.

However, my claim has been not only that we experience the results of our deliberate and straightforward mental agency as practically motivated; but also that we experience the straightforwardness of their motivation, meaning that we experience them as immediately responding to the respective practical reasons. To understand and support this thesis, it is helpful to consider first what it could mean to experience some mental episode as responding to practical reasons in a mediated way.

As already mentioned, mediated mental agency is characterized by the fact that it—often deliberately—relies on certain passive processes in order to bring about certain mental phenomena. For instance, when we act on the intention to finally force a conclusion on a certain matter in the light of the epistemic reasons already available to us, we usually do so with the expectation that the respective epistemic processes or mechanisms triggered by us are likely to compel us to endorse the proposition which best reflects our epistemic reasons. Now, our experience of successfully forming a judgement in this way shows two aspects, which correspond to two elements involved in such a formation. First, our initial attempt to come to a conclusion by setting in motion certain epistemic processes presents itself in phenomenal consciousness as active: we are aware of it as a rational response to our underlying desire or intention to force the issue. But second, the subsequently occurring impact of the triggered epistemic mechanisms presents itself to us as passive: we are aware of the compelling force of the epistemic reasons on our formation of the judgement and, more precisely, on our actually drawing one particular conclusion, rather than another. Accordingly, our complex experience of intentionally forming a judgement on the basis of evidence has a double character: it involves both an experience of the support provided by practical reasons and an experience of the influence of the epistemic mechanisms. And something similar will be true for other examples of mediated mental agency, whether they rely on epistemic

[29] In addition to O'Brien (cf. n. 28 above), Audi speaks of a 'phenomenal sense of acting in response' to some reason (1993: 154), Wegner of a 'feeling of voluntariness or doing a thing "on purpose" ' or of an '*experience* of consciously willing an action' (2004: 650), and Siegel of a 'special sense or experience of carrying out an intentional action' (2005: 280).

processes (as when we deliberately try to remember something) or on merely causal ones (as when we deliberately try to influence our mood by imagining something).

Because the effects of the passive processes deliberately triggered by us occur often almost immediately after we have begun (and finished) to perform the respective action, it might seem as if we experience a single mental phenomenon as both active and passive. But cases in which we fail to form a judgement despite all our attempts—say, because our evidence does not favour one conclusion over another and thus lets the epistemic mechanisms run idle—indicate that there are in fact two distinct phenomena with two distinct experiential characters. The mental action of setting in motion the epistemic processes (i.e. the attempt to judge the issue) is experienced as active, while the subsequent output of those processes (i.e. the judgement) is experienced as passive. Other cases, in which there is much more delay between the trigger and the product of the passive processes involved, or in which the triggering action is bodily, make this even clearer (cf. the example of inducing a judgement by deliberately taking a slow-acting drug).

Our experience of successfully forming a mental episode in a more straightforward way, on the other hand, does not show such a double character. Deliberately conjuring up an image will involve the awareness of the impact of practical motivation on the resulting image, but not the awareness of the impact of some epistemic or causal processes. Of course, we may sometimes become aware of some obstacles beyond our influence when attempting to perform a certain straightforward mental action. And we may experience their impact on us in a very similar way to how we experience the force of the epistemic or causal processes in the examples of deliberate mediated agency. For instance, when trying to visualize an object with twenty equal sides, we may realize that we cannot do this, and our attempt and recognition of failure may be accompanied by a strong feeling of the imposition of respective limits on our capacities involved. But if we succeed in forming the image, no such awareness of an obstacle or an external force will occur. Similarly, our choice of what to visualize may be influenced by epistemic considerations, and we may be consciously aware of this fact (e.g. when we decide to visualize a sunny forest partly because of concluding that this will calm us down). But this awareness of an epistemic impact will be part of our formation of the respective desire or intention to visualize. And it will therefore precede our straightforward agency of visualizing, as well as our experience of our engagement in this activity.[30]

[30] The case of guessing is equally unproblematic, but slightly more complicated, given that any potential impact of the evidence available to us need not precede our active choice of which proposition to accept, but instead may restrict it during our active engagement with it (cf. the experience of external objects as restricting our active bodily movement). In the light of what has been said above (cf. esp. n. 4 above), the agency constitutive of guessing counts as straightforward, despite the potential involvement of epistemic factors. For what we guess (in contrast to what we judge)

These considerations about the various ways in which we can deliberately influence what happens in our minds illustrate that our experiential awareness of mental agency seems indeed to be restricted to deliberate and straightforward agency and its mental products. But they equally link up to the observation that this kind of experience reflects especially the straightforwardness of the kind of agency concerned. For we experience the (not necessarily temporal) immediacy of determination with which the direct results of deliberate mental agency occur in response to the respective practical reasons. The idea is that we experience the mental episodes which we intentionally produce without exploiting certain passive processes as directly determined by and flowing from our motives and our attempts to act on them. And it is an essential part of this experience of immediacy that we are not aware of any determining factors other than practical motivation. In other words, we experience the immediacy of the practical motivation partly by not being conscious of any other determining factors as intervening between our desire or intention and the formed images, apart from our mental agency. When we visualize a sunny forest, we experience the resulting visual image as a direct response to our attempt to visualize a sunny forest and, given that this attempt flows immediately from our respective desire or intention, also as a direct response to the latter.

By contrast, the mental episodes produced in a mediated way by our deliberate mental activity do not present themselves as immediately responding to our practical motives, given that they are experienced as determined by epistemic or causal processes. When we intentionally form a judgement on the basis of the evidence available to us, we are aware of the compelling impact of the epistemic considerations determining which particular proposition we end up endorsing. And this aspect of our experience of the judgement is responsible for the fact that it cannot count as an experience of immediate practical motivation, given that this experience of immediacy requires the absence of any awareness of determining elements other than practical motives. This is precisely the reason why the phenomenal character of our judgements is incompatible with the phenomenal character of mental episodes resulting from our deliberate mental agency in a straightforward way. For independently of whether our judgements are actually motivated by epistemic reasons, they always phenomenally present themselves to us as such. And this experiential awareness of an epistemic rational influence would undermine—for the reason just mentioned—any awareness of an immediate motivational impact of practical reasons. But such an awareness of immediacy is always part of how we experience straightforwardly formed episodes. Hence, none of our mental episodes can be phenomenally marked

is ultimately a matter of our choice (cf. n. 4). Moreover, the related experience of straightforward agency is compatible with the potential simultaneous awareness of epistemic influence and limitation, given that the latter is never experienced as fully determining what is guessed (in contrast to what is judged), but as leaving room for the experience of the immediate impact of the practical reasons concerned.

for us both as a judgement and as a product of deliberate and straightforward mental agency. And this explains why we cannot form judgements at will: we would have to consciously experience the resulting episodes in a way which is not open to us.

The fact that our experience of the immediacy of the practical motivation of a given mental episode is incompatible with our simultaneous experience of the same episode as occurring in response to epistemic motivation is perhaps more fundamentally due to the fact that the respective two phenomenal aspects reflect incompatible features of our episodes. Since no episode can be both epistemically and straightforwardly practically motivated, it seems that, if one aspect of the experiential character of one of our episodes adequately reflects its epistemic motivation, another aspect cannot simultaneously adequately reflect the straightforward practical motivation of the episode. Similarly, that we cannot properly experience a perception both as representing red and as representing green is maybe primarily due, not to how such perceptions actually present themselves to us in phenomenal consciousness, but to the underlying fact that a red-perception cannot simultaneously (and with respect to the same part of an object) be a green-perception. Consequently, the incompatibility at issue may be located only derivatively in how we experience the respective mental episodes, and ultimately in which role these experiences play in our mental lives, namely to reveal the nature of the episodes concerned.[31]

Given that—as already noted before—our experiential awareness is fallible, it might however still be possible that we can actively form a judgement in direct and conscious response to some of our desires or intentions, as long as we do not experience the resulting judgement as a judgement, that is, as epistemically motivated.[32] But such a case would not count as an instance

[31] The preceding considerations apply equally well to the many mental phenomena, which mix imaginative or otherwise straightforwardly active elements with more passive—and often cognitive—elements. Deliberately trying to visualize a particular friend as sitting in the chair opposite to me will involve seeing the chair, actively recalling his appearance and imaginatively combining and manipulating the 'sensory material' thereby provided to conjure up the image of him sitting in the chair. And while the perceptual element occurs in a purely passive way, both the mnemonic and the imaginative element involve conscious agency, albeit the former in a mediated and the latter in a straightforward manner. What we thereby experience as active is precisely what we do straightforwardly: namely, whatever needs to be done to trigger the respective mnemonic process, as well as our conjuring up the image by using the provided 'sensory material'. And the same seems true of other more complex forms of mental agency. Calculating a sum in one's head, for instance, consists in actively triggering a series of epistemic processes (e.g. those providing us with the result of adding or multiplying two numbers). But although the impact of these processes is experienced as passive, we actively trigger them in a mediated way by means of performing a more basic straightforward action, coming with the respective experience of agency. And a last example is cases of visualizing where some of the details of the resulting image are passively 'filled in' by the mind (e.g. due to our knowledge or memory of generic appearances) and experienced as such, although the other aspects of the image are experienced as immediately determined by our imaginative agency.

[32] Thanks to Lucy O'Brien for pointing out this possibility.

of deliberately forming a judgement in a straightforward manner. We might have performed the described action on the basis of a desire or intention to form a supposition; and a mistake might then have led to the occurrence of a judgement experienced as a supposition. But then, we would have tried to form a supposition, and not a judgement. Alternatively, we might have intentionally set out to form a judgement in such a way that it is not experienced by us as a judgement. But then, we would have had to exploit some passive processes bringing about this phenomenal illusion and, hence, would not have formed the judgement in a straightforward manner. There is perhaps also the possibility that we might come up with and might act on the intention to form a judgement at will in such a way as to fail to experience it as the straightforward product of mental agency (i.e. as immediately practically motivated). But our action could not be successful, given that the satisfaction of the two intended goals—the straightforward formation of the judgement and the creation of the phenomenal illusion—dictates incompatible means. Since we cannot bring about phenomenal illusions at will, the achievement of the second goal requires the reliance on certain causal processes. But it is precisely such a form of mediated agency which is ruled out by the successful straightforward formation of the judgement.

It has also been argued that our phenomenal experience is systematically misleading with respect to the nature of our minds: either because there are no instances of judging, imagining, or deliberate mental agency, despite it seeming to us that way; or because there are such instances, but they are not as they seem to us to be (cf. the eliminativist approaches to phenomenal consciousness). Here is not the place to assess the respective arguments, but let me briefly note the consequences their soundness would have for the experiential approach to the involuntariness of judgements. If none of the mental phenomena at issue existed, the question of whether we can form judgements at will would not arise, but instead only the question of why it none the less seems to us as if there is a difference between judging and imagining in respect to (what merely appears to us to be) deliberateness and straightforwardness. The normative approach could not hope to answer this question about our phenomenology: if there were no judgements, then there would also be no norms for judgements. The experiential approach, on the other hand, would still have something to say about the difference between judging and imagining and would also have good chances to be compatible with the—presumably causal and subpersonal—account of why things erroneously seem to us a certain way in the first place. On the other hand, that our phenomenal experience might turn out to generally misrepresent certain aspects of the nature of our mental phenomena would not pose any problem for the experiential approach, as long as the latter remains true of how we actually do experience judgements and mental actions. If the phenomenal illusion concerned the seeming rational motivation of judgements or deliberate mental actions, this would in fact mean, again, that there are no judgements or deliberate mental

actions, given that it is essential to these mental phenomena that they are rational and normally rationally motivated. And if the phenomenal illusion concerned some other aspect of the nature of judgements or deliberate mental actions, this error would be irrelevant for the question of whether judgements can be formed at will. For instance, it might indeed be the case that our actions are caused by certain sub-personal factors in our minds, although we experience them as originating in our tryings or volitions, or in us as conscious agents (cf. Wegner 2004). But our awareness of deliberate and straightforward activity would still be correlated to the respective instances of mental agency; and it would still be incompatible with our awareness of judgements.

7. CONCLUSION AND OUTLOOK

What I have been trying to show is that the experiential approach succeeds in establishing two things: the psychological and non-normative constraint on all our possible judgements that they are always experienced by us as epistemically motivated; and the incompatibility of this constraint with their deliberate and straightforward formation and, in particular, with experiencing them as formed in such a way. My main conclusion is therefore indeed that, for us, judging cannot be active and deliberate in the same straightforward way in which imagining can be active and deliberate. But the preceding considerations have also further substantiated the idea that our conscious experience of agency is a guide to—and only to—deliberate and straightforward agency, at least if mental activity is concerned. If something mental is experienced as active, it is normally part of deliberate mental agency. Our awareness of mental passivity, on the other hand, seems less revelatory, given that it is still compatible with more mediated forms of deliberate agency, such as in the example of intentionally changing one's mood and, indeed, in cases of deliberately forming judgements on the basis of evidence.

The experiential approach is compatible with the idea that it is possible to conceive of judgements as being formed at will, and to desire or intend to form a particular judgement in this way. All it claims is that we are bound to fail if we consciously try to act on such a desire or intention. In this respect, it is likely to contradict the version of the normative approach which assumes that the truth norm (TN), and presumably also something like the constraint (C), are part of our concept of judgements, and that we have to employ this concept when aiming to form a judgement or classifying a mental episode as a judgement. Given the presupposition of certain further conceptual links (e.g. between truth and evidence), this assumption may be said—following an argument similar to the one presented during the discussion of (C)—to entail that we cannot conceive of judgements as formed at will. And this again seems to imply that we cannot deliberately produce them in a straightforward manner: either because we

cannot form the required desires or intentions in the first place; or because our necessary failure to conceive of judgements as judgements after their deliberate and straightforward formation would prevent us from acquiring the knowledge that we can perform this kind of action, while such knowledge appears to be necessary for deliberate agency.[33] But apart from the fact that the objections against the normative approach mentioned above also apply to this more complex version of it, the latter faces its own specific difficulties. Notably, it seems very doubtful that it can establish all the conceptual truths required; or, indeed, the claim that we (including children) have to possess and employ such a rich concept of judgements in order, say, to decide to make up our minds about a certain issue, or to desire forming a particular judgement (e.g. by some manipulative means) because it would make us happier if we did.

But the experiential approach has other advantages over the normative approach, in all its facets. Not only can it easily accommodate the deliberate formation of manipulated or induced judgements, it also promises to be extendable to non-normative involuntary mental episodes. The normative approach has nothing to say about why we cannot form, say, perceptions, sensations or feelings at will, given that these phenomena are not subject to norms or requirements similar to (TN) or (C). By contrast, all kinds of involuntary mental episodes are phenomenally conscious and thus permit, at least in principle, the application of an argument which concentrates on this feature of them. It may be argued, for instance, that the causal determination of perceptions or sensations by those aspects of the world or our bodies, which they inform us about, becomes salient in their phenomenal character; and that this aspect of how we experience them is, again, incompatible with experiencing the immediacy involved in deliberate and straightforward mental agency. The experiential approach may therefore allow for a much more unified account of the involuntariness to be found in our conscious mental lives than the normative approach.

If the experiential approach indeed turns out to be the right one, then the involuntariness of our judgements is a matter of our psychology, and not of our concepts: it depends first of all on how we, as a matter of fact, experience judgements and mental actions. This leaves room for the possibility that the involuntariness of our judgements is merely contingent, and that there might be other creatures who experience these mental phenomena in very different ways and, hence, may still be able to form judgements at will. So far, the experiential approach has said nothing about this possibility. But it might perhaps be supplemented in such a way as to rule it out and thus to ensure the necessity of our inability to form judgements at will. The idea would be that the rational aspect of the actual phenomenal character of our judgements and mental actions is essential not only to how we in fact experience these conscious

[33] This seems to come very close to the first argument against voluntary beliefs to be found in Williams (1973).

mental phenomena, but also to how any potential being having them would experience them. And this might perhaps be traced back to the idea that the underlying rational nature of the respective phenomena can be phenomenally revealed to subjects experiencing them solely in the way in which it is actually disclosed to us—say, because how we experience these phenomena is constitutive of, or constituted by, or otherwise inseparably linked to, how they really are. This would still allow for experiential differences among subjects both of the same and of distinct species, as long as they do not concern the phenomenal disclosure of their rational nature (or of the respective aspects thereof). But the latter would necessarily be salient to all subjects in the same way and, hence, give rise to the same phenomenal incompatibilities. This would explain why we experience judgements and mental actions the way we experience them: as of necessity, to experience some mental phenomenon as a judgement or as a mental action would just mean, partly, to experience it as occurring in response to reasons. However, it would perhaps also imply that our experiential awareness of the rational aspects of our conscious minds is not primitive, and that a more fundamental account of it and of why we cannot form judgements at will can be formulated—namely in terms of those of their features which are constitutively linked to their phenomenal character and ultimately make up their nature.[34]

REFERENCES

Armstrong, David (1993) *A Materialist Theory of Mind* (London: Routledge).
Audi, Robert (1993) *Action, Intention, and Reason* (Ithaca, NY: Cornell University Press).
Bennett, Jonathan (1990) 'Why is Belief Involuntary?', *Analysis*, 50: 87–107.
Bodrozic, Davor (ed.) (2004) 'Intuition and Epistemology', *Dialectica*, 58: 3–10.
Burge, Tyler (1996) 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society*, 96: 91–116.
Carruthers, Peter (2000) *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press).
Cruz, Joseph, and John Pollock (1999) *Contemporary Theories of Knowledge* (Lanham, MD: Rowman & Littlefield).

[34] Various versions of this chapter have been presented at a one-day conference on mental agency in Senate House, London, the research colloquium at the University of Fribourg, the SOPHA meeting 2006 in Aix-en-Provence, and a conference on the phenomenology of agency, again at the University of Fribourg. For very helpful comments on these occasions, I would like to thank Julien Deonna, Julien Dutant, Pascal Engel, Guy Longworth, Martine Nida-Rümelin, Chris Peacocke, Joelle Proust, Nishi Shah, Joel Smith, Gianfranco Soldati, Juan Suarez, Fabrice Terroni, Stephen White, and Ann Whittle. For reading previous drafts and providing extensive comments, I am extremely grateful to Davor Bodrozic, Adrian Haddock, Lucy O'Brien, Matthew Soteriou, Gian-Andri Toendury, and an anonymous referee. I would also like to thank the Swiss National Science Foundation for funding part of the work on this chapter.

Dokic, Jérôme (2003) 'The Sense of Ownership: An Analogy between Sensation and Action', in Naomi Eilan and Johannes Roessler (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press), 321–44.

Dorsch, Fabian (2009) *The Unity of Imagining* (Frankfurt and Lancaster: Ontos Publishing House).

——and Gianfranco Soldati (2004) 'Conceptual Qualia and Communication', http://projectsfinal.interaction-ivrea.it/2004-2005/SYMPOSIUM 2005/communication material/.

Dretske, Frederick (1999) 'The Mind's Awareness of Itself', *Philosophical Studies*, 95: 103–24.

——(2000) 'Norms, History, and the Constitution of the Mental', in *Perception, Knowledge and Belief* (Cambridge: Cambridge University Press), 242–58.

Engel, Pascal (2002) 'Volitionism and Voluntarism about Belief', *Croatian Journal of Philosophy*, 2: 267–79.

Hornsby, Jennifer (1980) *Actions* (London: Routledge).

——(2005) 'Knowledge and Action', unpublished.

Lycan, William G. (1996) *Consciousness and Experience* (Cambridge, MA: MIT Press).

Martin, M. G. F. (1993) 'The Rational Role of Experience', *Proceedings of the Aristotelian Society*, 93: 71–88.

——(1997) 'Self-Observation', *European Journal of Philosophy*, 5: 119–40.

——(1998) 'An Eye Directed Outward', in Cynthia MacDonald, Barry C. Smith, and Crispin Wright (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 99–121.

Noordhof, Paul (2001) 'Believe What You Want', *Proceedings of the Aristotelian Society*, 101: 247–65.

O'Brien, Lucy (2003) 'On Knowing One's Own Actions', in Naomi Eilan and Johannes Roessler (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press), 358–82.

O'Shaughnessy, Brian (1980) *The Will* (Cambridge: Cambridge University Press).

Owens, David (2000) *Reason Without Freedom* (London: Routledge).

——(2003) 'Does Belief Have an Aim?', *Philosophical Studies*, 115: 282–305.

Papineau, David (1999) 'Normativity and Judgement', *Supplement to the Proceedings of the Aristotelian Society*, 73: 16–43.

Peacocke, Christopher (1996) 'Entitlement, Self-Knowledge and Conceptual Redeployment', *Proceedings of the Aristotelian Society*, 96: 117–58.

——(1998) 'Conscious Attitudes, Attention, and Self-Knowledge', in Cynthia MacDonald, Barry C. Smith, and Crispin Wright (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 63–97.

——(2003) 'Action: Awareness, Ownership, and Knowledge', in Naomi Eilan and Johannes Roessler (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press), 94–110.

Pink, Thomas (1996) *The Psychology of Freedom* (Cambridge: Cambridge University Press).

Shah, Nishi (2003) 'How Truth Governs Belief', *Philosophical Review*, 112: 447–82.

——and David Velleman (2005) 'Doxastic Deliberation', *Philosophical Review*, 114: 497–534.

Shoemaker, Sidney (1994) 'Self-Knowledge and "Inner Sense"', *Philosophy and Phenomenological Research*, 54: 249–314.

Siegel, Susanna (2005) 'The Phenomenology of Efficacy', *Philosophical Topics*, 33: 265–84.

Siewert, Charles (1998) *The Significance of Consciousness* (Princeton: Princeton University Press).

Wedgwood, Ralph (2002) 'The Aim of Belief', *Philosophical Perspectives*, 16: 267–97.

Wegner, Daniel M. (2004) 'Précis of: The Illusion of Conscious Will', *Behavioral and Brain Sciences*, 27: 649–92.

Weatherson, Brian (2007) 'Deontology and Descartes' Demon', http://brian.weatherson.org/ddd.pdf.

Williams, Bernard (1973) 'Deciding to Believe', in *Problems of the Self* (Cambridge: Cambridge University Press), 136–51.

Williamson, Timothy (2000) *Knowledge and its Limits* (Oxford: Oxford University Press).

Winters, Barbara (1979) 'Believing at Will', *Journal of Philosophy*, 76: 243–56.

Wollheim, Richard (1984) *The Thread of Life* (New Haven, CT: Yale University Press).

Wright, Crispin (1998) 'Self-Knowledge: The Wittgensteinian Legacy', in Crispin Wright, Barry C. Smith, and Cynthia MacDonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 15–45.

4

Reason in Action

John Gibbons

There are two ways in to the notion of an action. First, we might start with the intuitive distinction between active and passive, find some clear cases on each side of the distinction, and go from there. When you throw a rock at a window, you're active, and the rock is passive. Unfortunately, the unaided intuitive distinction between the active and passive, assuming for the moment that there's just one such distinction, does not seem able to withstand much pressure. Compared to you, the rock seems passive. But think about the moment of impact when the rock breaks the window. Compared to the glass, the rock seems active. The thought that every action has an equal and opposite reaction may involve one intuitive distinction between the active and the passive, but it's not the one we're after when we think about normal human actions. On the other side, if you do your best to keep perfectly still because there's a burglar in the house, or you take a break after a long day at work, it's not clear that the intuitive notion will count these as activity. But there is a strong inclination to count them as normal human actions.

The philosophically more common way into the notion of an action is through the notion of a reason. You throw the rock, keep still, and take a break for reasons. At least in the easy cases, this is enough to convince us that you do these things on purpose, and this is enough to convince us that these things are actions. The hope is that these two ways come to basically the same thing. The idea is that we can, perhaps someday, give a philosophical account of the nature of action in terms of the notion of a reason, and this account will correspond to at least one fairly intuitive conception of what it is to be active. You're active in the relevant sense when you're moved in the right way by reasons. People can agree on this while disagreeing about a great many things. We might disagree about whether reasons are causes, about whether practical reasons require the presence of an independent desire, and about how to individuate events while agreeing on this basic idea.

I think the basic idea is right. But there is a problem, and it looks like a problem for the basic idea, not merely a problem for any particular implementation of the

idea. The problem stems from the fact that, even if reasons are causes, causation by practical reasons isn't sufficient for being an intentional action. After all, desires and intentions are caused by practical reasons that rationalize them, but they're clearly not actions. Even if all actions are events or changes and desires and intentions aren't, the acquisition of a desire or an intention is an event, but it isn't always an action. If we can't understand the nature of action in terms of causation by practical reasons, how should we understand it?

The problem only arises if you believe that the intentional action of going to the bank and the non-action of wanting to go to the bank are both caused by practical reasons that rationalize them. This is at least how it looks at first. You might want to go to the bank because you need some money and know that that's where it is. And you might go to the bank for the very same reasons. Furthermore, it looks as though the relation between the reasons and the desire is the same as the relation between the reasons and the action. This relation involves not only causation, but causing and making sense of.

But maybe things are not as they appear. Maybe the relation between reasons and actions is different from the relation between reasons and desires or intentions. It's natural to suppose that, when *A*-ing is an action, reasons for *A*-ing have to be about *A*-ing. If the mental state that causes you to go to the bank counts as a reason for going to the bank, then the notion of going to the bank has to figure in the content of the cause. You might believe that going to the bank is a means to an end, or think you should go to the bank, or just want to go to the bank. If reasons for *A*-ing are always about *A*-ing, but reasons for desiring or intending are only about the objects of desire or intention, if they're first-order rather than second-order, then maybe the relation between reasons and actions is different from the relation between reasons and desires or intentions. Maybe, when it comes to actions, the relation involves not just causing and making sense of, but also being about.

The problem with this solution is that there are mental actions. Under the appropriate circumstances, considering the proposition that p, judging that p, deciding to ϕ, and trying to ϕ are all things that you do. They're actions, and since these descriptions entail that you have a mental representation of p or of ϕ-ing, they're mental actions. But if reasons for *A*-ing were always about *A*-ing, and deciding to ϕ is an action, it would follow that reasons for deciding to ϕ are second-order, that they're about deciding. But they're not. Reasons for deciding to ϕ are first-order. They're about ϕ-ing. Or so I will argue. If the reasons for the actions of deciding or trying to ϕ are the same as the reasons for the non-actions of intending or desiring to ϕ, then the difference between actions and non-actions is not a difference in the kinds of reasons that cause or rationalize them. And the difference is not that the relation between reasons and actions involves causing, making sense of, and being about. So what is the difference?

This chapter is structured around two different ideas about practical reasons. The first we've seen already. This is the idea that reasons for *A*-ing are always

about *A*-ing. In the first section, I sketch an outline of a theory of action in terms of the notion of a reason that's based on this idea. I try not to rely very heavily on my answers to certain controversial questions about, for example, the role of desire in practical reason or the individuation of events, since I don't think the problem depends on any specific answers to these questions. The resulting view can correctly identify certain intuitively passive mental events as non-actions. But it cannot correctly identify certain intuitively active mental events as actions.

In the second section of the chapter, I reject the idea that reasons for *A*-ing are always about *A*-ing, and replace this with the idea that first-order propositional attitudes rationalize (or are reasons for) first-order propositional attitudes; second-order propositional attitudes rationalize second-order propositional attitudes; and so on. The resulting picture of action in terms of reasons can correctly identify our intuitively active mental events as actions. But it incorrectly identifies the intuitively passive mental events as actions as well. In the third and final section, I try to sketch yet another picture of the nature of action, still in terms of the notion of a reason, that gets the intuitive distinction right.

1. REASONS FOR A-ING ARE ABOUT A-ING

To make sense of my talk about reasons (call them 'motivating reasons' if you like), I'll begin with a picture of the rationality of actions. The rationality of an action is completely determined by (1) the rationality of the reasons for which you perform the action, i.e. the mental states that cause the action, and (2) the rationality of the move from those reasons to the action. It's a familiar idea that there's a distinction between reasons for and reasons for which.[1] You might have reasons for doing something but they are not the reasons for which you do it. It's a further step, but a step I'm willing to take, to say that, in general, the difference between reasons for and reasons for which is a causal difference.[2] The reasons for which you do it are the reasons that cause the action. And it seems that these are the only reasons relevant to the justification or rationality of the action. You get the same thing in the theoretical case. If you believe that p and that if p then q, but these have nothing to do with your believing that q, if your belief that q is the result of wishful thinking, then your belief that q is not justified.

Part of the study of practical reasoning is the study of which moves from reasons to action are rational. For example, the practical syllogism doesn't seem to care which way the arrow goes. If you believe that *A*-ing is a sufficient condition

[1] Donald Davidson, 'Actions, Reasons, and Causes', in *Essays on Actions and Events* (Oxford: Clarendon, 1980).

[2] Does the problem I'm raising crucially depend on this assumption? No. As we'll see, the problem arises for any view according to which the relation between reasons and actions is distinct from the relation between flipping the switch and turning on the light.

for *B* and you want *B*, it seems that these mental states could rationalize desiring to *A*, intending to *A*, deciding to *A*, trying to *A*, and intentionally *A*-ing. But if you believed that *A*-ing was a necessary condition for *B* and wanted *B*, these mental states could rationalize the same set of propositional attitudes. Of course, there's only an anemic sense in which the unfortunately named practical syllogism justifies actions and the rest. It gives you (2) but not (1). It may be a rational move, but that doesn't mean that you're moving from a reasonable place. You get the same thing in the theoretical case. If your belief that *p* and the belief that if *p* then *q* cause you to believe that *q*, this is a rational move to make. But if your belief that *p* is completely unjustified, then you're not justified in believing that *q*. In what follows, I'll be primarily concerned with (2) rather than (1). Given that a set of mental states causes a propositional attitude or action, what does it take for the set to rationalize or justify the attitude or action? In particular, we're interested in what justifies mental actions.

The first thing you want to say about mental action is something like this. Some of the stuff that goes on in your mind is stuff that happens to you. But at least some of what goes on is stuff that you do, mental action. I'd like my theory of action to apply as much to these cases as they do to the rest. But examples that illustrate the distinction seem easy enough to come by. You're reading the paper and a car alarm goes off in your neighborhood. Your auditory experience of the alarm and your coming to believe that there's a car alarm going off nearby are things that happen to you. The only thing you're doing, reading the paper, has nothing to do with these mental events. But when you're working on a philosophical argument, or trying to solve a puzzle, or doing mental arithmetic, these seem like things that you do. They seem like mental actions.

So consider this puzzle about planning. Planning is something you do, at least occasionally, rather than something that happens to you. Or so it seems. You have a number of things to do; you can't do them all at once; so you need to plan your day. Planning looks like a mental action. Like many actions, you plan by doing other things: you consider options; you evaluate them; you rank them or compare them with each other; and sometimes, you settle on one. If planning is an action, and you do it by doing these other things, these other things are actions as well.

Now what's an action again? Actions are things that happen for reasons. Since there are unintentional actions, we have to be careful here. Let's try this theory. Your *A*-ing is an action when it's intentional under some description.[3] Or, if you don't like that one, *x* is an action when it's on the same action tree as an intentional action.[4] The main difference between these views is how they individuate events. Both agree that calling it an action does not mean that it's

[3] See Davidson, 'Actions, Reasons, and Causes' and 'Agency' in *Essays on Actions and Events*. Also see G. E. M. Anscombe, *Intention* (Ithaca, NY: Cornell University Press, 1958).

[4] See Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970). For more on this way of individuating events, see Jaegwon Kim, 'Events as Property Exemplifications', in *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

intentional, or intentional under that description. The notion of an unintentional action is not contradictory. Still, you understand actions in general in terms of intentional ones: something's an action when it's intimately related to, either identical with or on the same action tree as, an intentional action. In what follows, I'll speak the first way. If you turn on the light by flipping the switch, your turning on the light just is your flipping the switch. Feel free to read that 'just is' as 'is on the same action tree as.' I promise, nothing will turn on this.

So what is it for an action to be intentional? This one's a little harder. You intentionally *A* roughly when you *A* for a reason for *A*-ing. Now we have two occurrences of the word 'for.' You *A* for a reason, and your reason, whatever it is, is a reason for *A*-ing. Since we're assuming that the difference between reasons for and reasons for which is a causal difference, it's natural to suppose that you *A* for a reason when the reason is causally related in the appropriate way to your *A*-ing. And we might say that a reason is for *A*-ing when it represents *A*-ing in a favorable light. This is the idea that reasons for *A*-ing are always about *A*-ing. If it represents *A*-ing in a favorable light, then it must represent *A*-ing. So it seems that you intentionally *A* only if you see *A*-ing as a means (in some suitably broad sense) to an end, or you see *A*-ing as an end in itself, or you think you ought to *A*.

I take it that all of this is familiar.[5] We have a set of motivated, necessary conditions for an action to be intentional, and we have a necessary condition for being an action. You're an action only if you're caused by mental states that rationalize some intimately related action, either the same action under another description, or a distinct action on the same action tree. Maybe intrinsically motivated basic actions are caused by desires alone or the relevant normative beliefs, but all other actions require causation by instrumental beliefs. Perhaps you want more, and this is understandable. But we have enough to raise some questions.

I promised you a puzzle about planning. So let's take our little theory and apply it to that case. You plan by, among other things, considering options. It seems that you consider those options in order to plan. Your considering those options has a point, or purpose, or a goal, namely, your planning your day. So it seems that, if considering an option is a mental action, it's an instrumental, intentional action.

So let's take a particular case of planning in which you consider going to the bank. Your considering going to the bank is, or is very much like, entertaining a proposition, perhaps the proposition that you will go to the bank. I'm not too concerned with the kind of content we have here. It could be entertaining, and eventually evaluating, an action type, a part of a proposition if you like. I'm more

[5] For more of the details see e.g. Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980); Myles Brand, *Intending and Acting* (Cambridge, MA: MIT Press, 1984); Alfred Mele, *Springs of Action* (Oxford: Oxford University Press, 1992) and Berent Enc, *How We Act* (Oxford: Oxford University Press, 2003).

concerned with the attitude, and the attitude of considering or entertaining is just a matter of having it in mind, whatever the it may be.[6] So now we can put our question like this. Is entertaining a proposition ever an intentional action? And the answer appears to be no.

According to our little theory, if entertaining the proposition that *p*, or considering going to the bank, are intentional actions, they must be caused by beliefs or desires that include that description of the action in their content. If considering going to the bank is an instrumental intentional action done for the sake of planning, then it must be caused by a belief that considering going to the bank is a means to planning. But look at the content of that belief. If you have in mind the proposition *considering going to the bank is a means to an end,* then you have *going to the bank* in mind. You can believe that if *p* then *q* without believing that *p*, but you can't occurrently believe that without having the proposition that *p* in mind. If the proposition has conceptual constituents, you can't have the proposition in mind without having the constituents in mind. So the occurrence of the belief about considering going to the bank entails that you are already, or thereby, considering going to the bank, and so the belief can't cause the considering. We'll see this pattern again.

I asked if entertaining a proposition could ever be an intentional action. If it's an instrumental action, it has to be caused by a belief with the right sort of content. But the occurrence of the belief presupposes the entertaining, and so can't cause it. Maybe entertaining a proposition could be an intrinsically motivated, basic action. If it's a basic action, you don't need beliefs about how to do it, and if it's intrinsically motivated, i.e. done for its own sake, you don't need beliefs about what it will get you. But the issue is not specifically about instrumental beliefs. According to our little theory, intrinsically motivated basic actions are caused by desires or normative beliefs. But if you occurrently want to think about going to the bank, you are thereby thinking about going to the bank. If you think you should think about going to the bank, you are thinking about going to the bank. And if you intend to think about going to the bank, that's just another way of thinking about going.

In order to run the objection, we don't need to assume that your thinking about going to the bank in these circumstances is an action, and we don't need to assume that it's not. If it is an action, it's not an action in virtue of being caused by practical reasons. This is a problem for the theory I'm going after. If it's not an action in this case, it's hard to see how it could be in some other case, at least as long as we think about action in terms of causation by practical reasons. I'm not suggesting that this is a problem for every conceivable theory of action. I'm

[6] I'm not suggesting that this is the only thing 'considering' could mean. This is just considering in the philosophical sense. Perhaps in the ordinary sense, considering involves taking possibilities seriously. But as we'll see, I think the same puzzle I'm raising applies to taking possibilities seriously. But that will have to wait. For now, we're just concerned with having possibilities in mind.

just going after the one I find most plausible. This is the theory I sketched at the beginning of this section, a theory based on the idea of causation (or something like causation) by practical reasons.

For any case of ordinary, overt action that I can think of, if you intentionally *A*, you first have *A*-ing in mind. There may be a distinction between intending and foreseeing, and it may be that some of your merely foreseen actions are intentional.[7] But if you foresee *A*-ing, then you have *A*-ing in mind before you do it. Maybe you don't need an intention to *A* for your *A*-ing to be intentional. Maybe the intention to try to *A* is enough.[8] But you can't intend to try to *A* without having *A*-ing in mind. Our little theory is based on the intuition that the distinction between what you do intentionally and everything else you do is at least partly determined by what you have in mind. But if you replace '*A*-ing' with a description of a mental action like entertaining a proposition or considering going to the bank, you can't, logically can't, have *A*-ing in mind without thereby performing the action.

One reaction to what I'm calling a puzzle is to accept the conclusion. Right, entertaining the proposition that *p* or considering going to the bank are never intentional actions. One problem with this reaction is that it leaves curious the status of the claim that your considering going to the bank is done in order to plan. It's not too far from here to the claim that your considering going to the bank is done for a reason. And you have practical reasons not only to consider some possibility or another. You have practical reasons to consider going to the bank. I'm concerned with considering because it's a step you take in the process of planning. Unless the entire process of planning your day is a basic action, you do it by doing other things, by performing other actions. And if those other things are actions, we need to know under what descriptions they are intentional. I'm concerned with planning because it's a special case of reasoning, call it practical reasoning. And I'm concerned with reasoning because it's the paradigm of intentional mental action. Of course, thoughts, beliefs, desires, and intentions sometimes just come to us. If you think of control in terms of intentional action, the question is whether even occasional control over the contents of our own minds is an illusion.

So, back to the puzzle. Once you see the pattern here, you can find it in a number of places. Believing that entertaining the proposition that *p* is a means to an end can't cause you to entertain the proposition that *p*. It just is entertaining that proposition. Again, you can read the 'just is' as 'is on the same action tree as.' The pattern is clearest in this case because having a belief with that content entails entertaining the relevant proposition. But the 'just is' we're interested in is about tokens, not types. There's no entailment between the types *flipping a switch* and *turning on a light.* You can flip a switch to turn off a light or to turn on a computer.

[7] Gilbert Harman, 'Practical Reasoning', *Review of Metaphysics,* 79 (1976): 431–63.
[8] Michael Bratman, 'Two Faces of Intention', *Philosophical Review,* 93 (1984): 375–405.

Since it matters, I'll risk belaboring the point. What matters is the following distinction. On the one hand, there's the relation between flipping the switch and turning on the light, and on the other, there's the relation between turning on the light and walking across the room. When you turn on the light by flipping the switch, once you've flipped, there's nothing further you need to do in order to turn on the light. There's some sense in which you don't first do one thing and then do another. There has to be just one something here, either one action or one action tree. I use 'just is' to express this relation, whatever that may be. Whatever your view on the nature of this relation, your flipping the switch may cause the light to go on, but it doesn't cause your turning on the light. On the other hand, you might turn on the light in order to walk across the room. But here, once you've turned on the light, there is something further you need to do. Here, first you turn on the light, and then you walk across the room. Whatever the somethings are, either actions or trees, here you've got two. And maybe, your turning on the light is a partial cause of your walking across the room.

There's more to planning than merely having options in mind. You need to evaluate them. This involves, among other things, raising questions about the option. So suppose you ask yourself about the likely consequences of going to the bank, or you ask yourself what the downside is to going to the bank. Are these intentional actions? Well, first, can the relevant sorts of beliefs cause them? It seems not. When you're planning, realizing that asking yourself about the consequences is a means to planning just is taking that question seriously. And taking that question seriously just is asking yourself. In most cases when you *A* in order to *B*, you have to first see the connection between *A*-ing and *B*-ing. But in these cases, seeing the connection between *A*-ing and *B*-ing just is *A*-ing. It's not that once you see the relevance of the question you have to go on and do something further in order to ask yourself that question. This is more like flipping the switch and turning on the light than it is like turning on the light and walking across the room.[9]

Like the case of considering, nothing turns on focusing exclusively on instrumental beliefs. In the right circumstances, occurrently wanting to know about the downside just is a way of raising the question, and if there are proximal intentions, intentions about what to do right now, the same seems to go for them. Unlike the case of considering, there may be no entailment between the types. You can want, intend, or think you ought to figure out free will someday without taking the question seriously right now. But this is like saying that you can flip a switch without turning on a light. We're examining a view about tokens, about what's what in the circumstances. On my preferred way of speaking, there's just one mental event there. It just is your seeing the relevance of the question; it just is your seeing the connection between raising the question and planning; and

[9] In much the same way, realizing that in these circumstances, taking the possibility that *p* seriously is the thing to do just is taking that possibility seriously.

it just is your raising the question. But whether the relation is identity or level generation, it's not causation.

Two more examples, then we'll try to figure out what's going on. Sometimes intentions just come to us. You open the fridge, looking for something to eat. You notice the leftover pasta, and straightaway you act. The intention you act on, assuming there is one, is something that happens to you. In other cases, we form an intention or we make a decision, and these look like things that we do. Is forming an intention to *A* or making a decision to *A* ever an intentional action? You have to choose between *A* and *B* since you know you can't do both. All along, you dispositionally believe that forming an intention to *A* is one way of making up your mind, and you believe that forming an intention to *B* is also a way of making up your mind. But at some point you come to believe that forming the intention to *A* is a better way of making up your mind. But this, in the circumstances, just is forming the intention to *A*.

According to our theory, if forming the intention to *A* is itself an intentional action, you need an instrumental or normative belief, or a desire, intention, or what have you with the appropriate content to cause the intention. When you have the same attitude toward the two options, intending to *A* and intending to *B*, neither of the instrumental beliefs can cause an intention. If you want a cause that will explain why you formed the intention to *A* rather than *B*, then it has to be your taking an attitude toward one of the options that you don't take toward the other. Maybe it's coming to prefer an option, or realizing that you prefer an option, or coming to see an option as better. But whatever the attitude is, your taking that attitude toward intending to *A* just is your taking that attitude toward *A*-ing. We're looking for the right kind of cause of your intending to *A*, but all we find is the intention to *A* under another description.

I'm sure you can think of cases in which someone thinks it would be better to intend to A and yet fails to intend to A. These are just like cases in which someone thinks it would be best to *A* and yet fails to *A*, and even fails to intend to *A*. I'm not trying to give an analysis of what it is for someone to settle on, decide on, or opt for something. I'm not concerned with the relations between evaluative judgments and intentions. I'm concerned with the relation between settling on intending to *A* and settling on *A*-ing, whatever settling on may be. And even here, I'm concerned with the relation between tokens of these types, not the types themselves. According to our theory, in order for your decision to *A* to be an intentional action, it needs to be caused by thoughts about deciding or intending to *A*. At least in ordinary circumstances, once you've settled on intending to *A*, you have already settled on *A*-ing.

Again, we might consider accepting the conclusion that deciding to *A* is never an intentional action. This goes against my intuitions, but maybe people differ about this, and anyway, we're all familiar with the idea that, when the going gets rough, some intuitions must be sacrificed in order to save others. So maybe we should try out the idea that the action is intentional under the description

'making some decision or another' rather than under the description 'deciding to *A*.' I have two problems with this. First, we ask if the reasons can cause the action. You know that if you make some decision or another you can get on with your day, and you want to get on with your day. Is this reason enough to make some decision or another? Presumably not. Your reasons for making some decision or another don't choose between your options. But making some decision or another just is choosing between your options. So no matter how strongly you feel the need to decide, if you don't have some reason to decide to *A* or some reason to decide to *B*, you'll remain undecided. So unlike reasons for going to the bank, reasons for making some decision or another are never enough to get you to do the thing in question.

The second problem with the suggestion may be the flip side of the first. In the usual case, you don't just have reasons for making some decision or another. You have reasons for deciding to *A*. And given the reasons for deciding to *A*, reasons for making some decision or another aren't just insufficient. And they aren't just unnecessary. It's difficult to see how they help at all. So if, in the normal case, you have reasons for deciding to *A*, and you decide for those reasons, then we should be reluctant to give up on the idea that deciding to *A* is an intentional action. Or at least, we should be reluctant if we think there's an intimate connection between being an action and being moved by practical reasons.

Our intuitions, or perhaps, our philosophical views, might make the following curious distinction. Deciding to *A*, at least sometimes, is something we do. But coming to believe, or making a judgment, is always and everywhere something that happens to us. Put the view another way. Settling on a plan, that's up to us. Settling on a view, that's completely different. I'm not absolutely certain how we should treat these cases, but I do think we should treat them the same way. In any case, just to round out our set of examples, let's look at making a judgment and coming to believe.

Now, we're not concerned with cases like hearing the car alarm, cases that don't even look like actions. But some cases of coming to believe do look like actions. You want to know whether or not *p*. This desire causes you to initiate a process of inquiry. Maybe you just think about it, or you look something up, or you dirty your hands in the empirical world. The process of inquiry, something you do, results in your coming to believe that *p*. If something you do results in the light's going on, then you turn on the light, and this is something else that you do. If something you do results in your believing that *p*, then coming to believe that *p* is something you do. Our question is whether this is ever an intentional action, and our little theory seems to say that it's not. Here, I think, many people would be happy with the conclusion.

If the motivating intention or desire is to know whether or not *p*, then during the process of inquiry, you don't have, even dispositionally, either instrumental belief. You don't think that believing that *p* is a way of *knowing* whether or not *p*, unless you already believe that *p*. But maybe you just want to have a view. If

that's the desire, then this case parallels the case of forming an intention. You might dispositionally believe that coming to believe that *p* is a way of making up your mind, and you may believe this about coming to believe that not-*p* as well. We're assuming that whether or not coming to believe that *p* is intentional depends on its being caused by mental states with the appropriate content. If the belief is caused by the evidence itself, so to speak, it's not intentional. But if you do come to have the contrastive means–end belief, it's probably not a cause. Your coming to believe that your coming to believe that *p* is a better way of making up your mind just is your coming to believe that *p*.

Now there may be cases of belief acquisition that do count as intentional action according to our theory. You might think that coming to believe that *p* will make you feel better and want to feel better. These might cause, in the appropriate way, your coming to believe that *p*. The question is what it takes according to our theory for you to intentionally come to believe that *p*. Our theory appears to say that if the belief is motivated by the desire to feel better, it can be an intentional action. But if the belief is motivated by a desire for knowledge, then it can't. When you see coming to believe that *p* as a means to knowledge, you thereby come to believe that *p*.

The argument generalizes to judging in general. Judging, unlike believing, is occurrent. Unlike coming to believe, you sometimes judge things you've believed for some time. So in the course of your inquiry, you judge that *q*, or you occurrently believe that *q*, where this is something you've believed for some time. Is this, at least, an intentional action? I'm sure you can see what's coming. Seeing the relevance of judging that *q* to the course of your inquiry just is judging that *q*. If the belief's becoming occurrent is caused by other beliefs about the world, this might not even be an action, let alone intentional. If you do have an instrumental belief about judging that *q*, this is the same mental event under another description.

2. N-ORDER ATTITUDES RATIONALIZE N-ORDER ATTITUDES

What's gone wrong? We have a perfectly natural picture of reasons for action according to which reasons for *A*-ing have to be about *A*-ing. We apply this to the case of mental action, and it turns out that, even if we do have the relevant second-order beliefs or desires, beliefs or desires about considering, deciding, and the rest, they don't help. The problem is that we're asking second-order propositional attitudes to directly rationalize first-order propositional attitudes. But in general, n-order propositional attitudes only rationalize n-order propositional attitudes.

Part of the study of reasoning is the study of which moves it's rational for the mind to make. Supposing that a set of propositional attitudes, the reasons, cause another, the conclusion, under what conditions is this a rational transition? Two

sorts of considerations seem obviously relevant, the relation between the contents of the reasons and the content of the conclusion, and the relation between the attitudes of the reasons and the attitude of the conclusion. A belief that *A*-ing will lead to *B* and a desire to *B* can rationalize the desire or intention to *A*, but they can't rationalize the desire to *C* or the belief that you will *A*. The problem in the first case is that you don't have the right relation between the contents, and in the second, you don't have the right relation between the attitudes.

The rule, that n-order propositional attitudes only rationalize n-order propositional attitudes, holds for reasoning in general, both practical and theoretical, and its motivation comes from reflection on the right relation between the contents. Of course, there's much more to that relation than is embodied in this simple rule. The rule doesn't even explain what's wrong with the belief that *A*-ing will lead to *B* by itself causing a desire to *C*. But we know what's wrong with this case. The reasons are about one thing and the conclusion about something else. And this is what's wrong with violations of the rule about orders. A first-order propositional attitude is one whose content does not entail that you have any specific propositional attitudes, and a second-order propositional attitude is one whose content attributes to you a first-order propositional attitude. In short, first-order propositional attitudes are about one thing, the world, while second-order attitudes are about something else, your mind.

Now I think that, in every sense that matters, knowing that *p*, perceiving that *p*, and intentionally making it the case that *p*, or intentionally *A*-ing, are propositional attitudes.[10] Being factive does not keep you from being mental. After all, there's a clear sense in which your belief that you exist and your belief that two plus two is four are factive. There's no possible world in which the proposition expressed by the belief sentence is true and in which the proposition expressed by the content sentence is false. There are some clearly mental states that are factive, so being non-factive is not a defining feature of the mental. But instead of asking for defining features of the mental, I'll just ask you to recognize some sense in which intentionally *A*-ing is a first-order propositional attitude. Intentionally *A*-ing clearly has something important in common with desiring and intending to *A*. They all require a concept of *A*-ing. Failures of substitutivity in the relevant sentential contexts are all explained in the same way. And, most importantly for our purposes, the same sorts of mental states rationalize them all.[11]

[10] For the idea that knowledge is a mental state, see Timothy Williamson, 'Is Knowing a State of Mind?', *Mind,* 104 (1995): 533–65; John McDowell, 'Knowledge and the Internal', *Philosophy and Phenomenological Research,* 55 (1995): 877–93; John Gibbons, 'Knowledge in Action', *Philosophy and Phenomenological Research,* 62 (2001): 579–600.

[11] Maybe the problem with thinking of intentionally making it the case that *p* as a propositional attitude has nothing to do with factivity. Maybe nothing could be both an action and an attitude. When you raise your arm and your arm goes up, this is the sign of action, not attitude. (I'd like to thank an anonymous referee for Oxford University Press for this suggestion.) But again, reflection on mental action causes trouble for this hypothesis. Imagining that *p*, judging that *p*, deciding and

As long as we remain focused on deciding, judging, and the rest, we don't have to worry about factivity. Deciding to sink the *Bismarck* is a mental action on anyone's account. Since the proposition that you sink the *Bismarck* does not entail that you have any specific propositional attitudes—you could have done it unintentionally and so without any representation of sinking or of the *Bismarck*—any attitude you have toward this proposition (deciding, desiring, intending, or trying) is a first-order propositional attitude. To the extent that we accept the rule about orders, we should expect first-order propositional attitudes to rationalize all of these, actions and non-actions alike.

The rule about orders seems like a special case of the idea that the reasons and the conclusion have to be about the same things, and a difference in orders is a difference in what the attitudes are about. But aren't there cases where second-order considerations are relevant to first-order propositional attitudes? Of course there are. But none of the cases I've seen so far involve violations of the rule. I'll look at two such cases and show that, rather than violating the rule, they depend on it. This doesn't mean that there are no counterexamples. It just exhibits the strategy for dealing with apparent counterexamples.

Here are two stories. In the first case, you want to like reading fancy books (Proust and such) because you see something good about fancy books. Reading them will make you smarter. This looks like a second-order desire, the desire to desire to read fancy books rationalized by a first-order desire, the desire to be smart, plus some relevant beliefs. In the second case, you want to like fancy books because you want to be like your friends who like fancy books, and this leads to your reading them. This looks like a first-order action or intention to read fancy books rationalized by a second-order desire to like them.

What's going on in these cases? As the stories were told, a lot has been left implicit. When we make the reasoning explicit, we can find a ladder up or a ladder down. In the case of a propositional attitude with a complex content, say, a belief in a conditional, the belief could be first-order with respect to its antecedent and second-order or higher with respect to its consequent. In a simple case of practical reasoning, the belief that if you A this will lead to X and the desire for X rationalize the intention to A. It makes no difference what X is, as long as the content of the consequent matches the content of the desire. X could be the desire to intend to believe that p. As far as the rule about orders goes, all that matters is that the order of the antecedent is the same as the order of the conclusion, and the order of the consequent is the same as the order of the desire. Since the belief is first-order with respect to the means and fourth-order with respect to the end, it acts as a ladder down. It lets a fourth-order desire rationalize a first-order intention. This is what's going on in the cases described above.

trying to make it the case that p all look like propositional attitudes. And it certainly seems possible for some of them to be actions sometimes. Is there really any reason to think they can't be both?

In the first case, you want to like reading fancy books because you see something good about fancy books: you think reading them will make you smart. What does this mean? It looks like it means something like the following. You believe that if you read fancy books then you'll be smart, and you desire to be smart. But this belief–desire pair only rationalizes the desire, intention, or intentional action of reading fancy books. It only rationalizes something first-order. If this were all you cared about, it wouldn't make any difference to you what mental states you had. But it's part of the story that you want to like reading fancy books. Why is that? Presumably because you believe that you'll read fancy books only if you like them. And this is our ladder up. A belief that you'll read fancy books only if you like them plus our already derived desire to read fancy books rationalizes a desire to desire or like them.

The other case is similar, except that it involves a ladder down. You want to like fancy books because you want to be like your friends, and this somehow gets you to read them. How should we understand this? Well, you desire that if your friends desire fancy books then you desire fancy books and you believe that your friends desire fancy books. These rationalize the second-order desire to desire fancy books. So far, you might not be the least bit attracted to fancy books. You might only be attracted to the desire. But how could this get you to read them? Well, if you believe that if you read them then you'll come to like them, and you want to like them, this rationalizes reading and wanting or intending to read them.

If we only look at ordinary, overt action, it looks as though reasons for *A*-ing always have to be about *A*-ing. But when we apply this to the case of mental action, it doesn't work. When we look at practical reasoning and the rules about the right relations between the contents, it looks as though n-order propositional attitudes only rationalize n-order propositional attitudes, because something in the rationalizing attitudes has to be about what the rationalized attitude is about. As long as we think of intentionally *A*-ing as a first-order propositional attitude, there's no conflict here. Our original thought about ordinary, overt action is just a special case of the rule about orders. Where intentionally *A*-ing is a first-order propositional attitude, reasons for this have to be first-order. They have to be about *A*-ing. So when we ask about reasons for deciding to ϕ, we shouldn't apply the rule that only holds in special cases, that reasons for *A*-ing are always about *A*-ing. We should apply the more general rule. Since deciding to ϕ is first-order, so are its reasons. But this means that we can't think of the relation between reasons and actions in terms of causing, rationalizing, and being about. Nothing could both rationalize and be about the decision to ϕ. If it rationalizes the decision, it's first-order. If it's about the decision, it's second-order.

Given this way of looking at things, we can accept the natural idea that any reason for *A*-ing is also a reason for deciding to *A* and acquiring the intention to *A*. When we thought that reasons for *A*-ing had to be about *A*-ing, we couldn't

make sense of the notion of an actively acquired intention since second-order rationalizers presupposed the action and so couldn't cause it. But if your reason for deciding to *A* is that *A*-ing will lead to *B*, then maybe this first-order rationalizer doesn't presuppose the action, so maybe it can cause the action. But explaining actively acquired intentions in this way comes at a price. It's now quite difficult to explain passively acquired intentions. Of course, if the intention is the result of a blow to the head, or neurosurgery, or someone's slipping you an intention-to-*A* pill, these will all be passively acquired intentions. But the overwhelming majority of actual passively acquired intentions and beliefs are the result of what we now take as reasons for acquiring them. If your intention to eat the pasta is caused by practical reasons that rationalize it, why isn't acquiring that intention an action?

So we're back to our original problem. You have a set of mental events, the acquisitions of intentions. Only a proper subset of these events contains mental actions. The question about the nature of action is a question about what feature all and only actions have in common. Of course, this is an old-fashioned kind of question, and some people will be embarrassed to be caught asking it. But many people who talk about actions think they have an answer to this question, an answer in terms of rational causation. This is the little theory of action that I presented at the beginning of the chapter. If you believe some version of this theory, you shouldn't be concerned about asking the question. You should be concerned that your answer to the question is false. If intention acquisitions outside the proper subset are as much the result of rational causation as the members of the proper subset, then rational causation does not make the difference between actions and non-actions.

We get the same structure for all of our examples. There's the set of mental events with propositional content. All of these count as having a proposition in mind, but only some of them are actions. Not all occurrent beliefs or all cases of taking a question seriously are mental actions. We can, if we like, name our proper subsets. We can use 'deciding' so that it only applies to the active acquisition of intention. We can restrict our use of 'judging,' 'considering,' and 'raising a question' so that they only apply to actions. But at most this only helps us raise the question. It doesn't help us answer it. Now we want to know what deciding, judging, and the rest, so understood, have in common with all and only other actions.

There may be a short and somewhat helpful answer to this question. Deciding, judging, and the rest are all parts of the process of reasoning, and there's more to reasoning than rational causation. Since I think of intentionally *A*-ing as another propositional attitude, I have no trouble with the idea that sometimes an ordinary overt action could be the conclusion of practical reasoning. Our answer is somewhat helpful because it allows us to raise our question in its fullest

generality. Something's an action when it's a step in the process of reasoning.[12] Once you know what reasoning is, you know what an action is. But it is only somewhat helpful because it doesn't really answer our question. Now we have a set of mental state transitions, all of which are in accord with your favorite rules about reasoning. This set includes the causation of your intention to eat the pasta by your beliefs and desires. It also includes the causation of your belief about the car alarm by your auditory experience in the context of your background beliefs. But only a proper subset of these transitions involves mental action. The rest merely happen to you. We can restrict the word 'reasoning' so that reasoning is always something you do and never something that happens to you. But this is just a label for our proper subset. It doesn't tell us what all and only actions have in common. Again, this lets us raise the question but doesn't answer it. Now our question is this. What's the difference between reasoning and rational causation?

3. REASON IN ACTION

We understand the notion of an action in terms of the notion of an intentional action. And we understand the notion of an intentional action in terms of the notion of a reason. You intentionally *A* when you *A* for a reason for *A*-ing. This gives us two very interesting notions to work with. First, there's doing something for a reason. So far, we've been assuming that this happens when the reasons are causally related in the appropriate way to the action. Second, there's the notion of one thing being a reason for another. This is the notion we've been trying to figure out. When we thought that reasons for *A*-ing had to be about *A*-ing, we could explain why going to the bank because you need money is an action while wanting to go to the bank for the very same reason is not. A fuller specification of the reasons will reveal that they're about going to the bank, but they're not about wanting.

We should reject the idea that reasons for *A*-ing are always about *A*-ing because it entails that reasons for considering, judging, deciding, and trying are

[12] Hold on. Some mental action involves reasoning. But does all of it? What about imagining that five bed, four bath house in the suburbs, or fantasizing about the slate tile on the walls and floor of the spacious master bathroom? Should we think of imagining, fantasizing, and trips down memory lane as reasoning? Yes. There are two sorts of trips down memory lane. Some of these trips are aimless. One thought leads to another, and you're not in control of where you go. These are mental events, but they don't look like mental actions. But sometimes you think about a particular subject matter on purpose. The thinking itself need not involve any calculating. It might just be daydreaming. But I can't help but think that if you daydream on purpose, you daydream for a reason. Doing it for the fun of it or because you feel like it counts as doing it for a reason. If you do something for a reason, then the move from the reasons to the doing is reasoning, the last step of practical reasoning. I think this is as true of trips to the bank as it is of trips down memory lane. But you should at least accept it in those cases you feel comfortable calling mental.

second-order. If you had the second-order reasons, they couldn't in general cause the actions, so they couldn't be the reasons for which you act. And at least so far, the second-order reasons seem rationally superfluous since deciding and the rest can be made reasonable by your first-order states. But if we do give up the idea that reasons for *A*-ing have to be about *A*-ing, we can no longer explain why going to the bank because you need money is an action while wanting to go to the bank for the very same reason is not. Maybe it's time to look at that other very interesting notion, the notion of doing something for a reason. Maybe there's a difference between being moved by a reason and acting on one. And maybe, rational causation is enough for the former, but something in addition is required for the latter.

Whether you take your inspiration from Kant or Lewis Carroll, you need a Faculty of Reason to move the mind. In Carroll's story,[13] the Tortoise accepted the proposition that *p* and the proposition that if *p* then *q*, and he asked Achilles why he must or why he should accept the proposition that *q*. Achilles started off on the right track. He replied, in effect, that the Tortoise should believe that *q* because things he believes entail that *q*. But the Tortoise notoriously accepts this non-normative fact about the shape of logical space. He wants Achilles to derive an ought from an is, and no mere fact about what's true in the same worlds as what will amount for the Tortoise as an answer to the normative question. But the Tortoise only accepts part of what Achilles said. Achilles said that he *should* believe because this entails that, and the Tortoise only accepts the fact of entailment. He sees what they entail but he doesn't see them as reasons. If you're trying to get from the is the Tortoise accepts to the ought Achilles wants, no further logical truths will help.

According to one common way of telling the story, the moral is that you must distinguish between believing a proposition and accepting a rule. The Tortoise believes the proposition that this entails that, but he doesn't accept, I guess, Modus Ponens. Unfortunately, you don't always get quite as much clarification of this notion of accepting a rule as one might hope. It's got to be a little like a desire since it moves the mind. But you evaluate it like a belief. Modus Ponens might not be true, but it is truth preserving, and the rationality of accepting it doesn't seem to depend on what other desires you happen to have.

In any case, there's something odd about the idea that the rule the Tortoise needs is Modus Ponens. Modus Ponens, whatever it is, only makes reference to the propositions that *p*, *q*, and if *p* then *q*. It's first-order all the way through. If it does have any bearing on mental states, it should have the same bearing on all mental states with the relevant contents. So suppose you believe that *p* and merely hope that if *p* then *q*. If these cause you to believe that *q*, this is a kind of wishful thinking despite the contents of the mental states being in accord with Modus Ponens. Suppose you desire *E* and merely hope that *A*-ing will get you

[13] Lewis Carroll, 'What the Tortoise Said to Achilles', *Mind,* 4 (1895): 278–80.

E. If these cause you to *A*, this is a kind of wishful acting, despite the similarity in content to genuine cases of practical reasoning. It looks like the rules have to be second-order. But it's not clear how to put them. We might try this. If you believe this and you believe that then you should believe the other. This is just the ought Achilles was after. If that's what you mean by a rule, there's no difference between my way of telling the story and the common way.

There is, of course, one thing we know about rules other than the fact that accepting one is not believing a proposition. There's a difference between following a rule and merely acting in accord with a rule. Suppose the rule is to take an umbrella if it's raining. On the basis of the weather report you saw last night, you think it's a perfect day for a walk. On your way out the door, you reach for your walking stick and grab your umbrella by mistake. Once outside, you see rain in the sky and an umbrella in your hand. You've acted in accord with the rule, but you haven't followed it. Perhaps the difference between reasoning and rational causation is the difference between following and acting in accord with the rules about reasoning.

So what is it to follow a rule? It looks like there are the following three necessary conditions. If the rule is to *A* in condition *C*, then you have to believe that *C*; you have to intentionally *A*; and you have to take your belief that *C* as a reason to *A*. Using our umbrella rule as an example, if you don't believe it's raining, you're not following the rule. If you don't take your umbrella on purpose, you're not following the rule. And even if you know it's raining and take your umbrella on purpose, if your belief that it's raining is not one of the reasons for which you take your umbrella, if you take it for some other reason, then you're not following this rule. Maybe these conditions are jointly sufficient. Maybe not. But let's look at what we have.

There are two ways of understanding our third condition, that you take *C* as a reason to *A*. On the weaker reading, your belief that *C* has to be one of the reasons for which you *A*. Roughly, your belief that *C* has to be causally related in the appropriate way to your *A*-ing. The notion of a reason need not enter the content of any of your thoughts. This notion of following a rule gives you at most rational causation. If the belief that *C* and the other first-order mental states that cause your *A*-ing are conceptually or rationally connected to *A*-ing, then you've got rational causation. This sort of thing happens in all our cases of passively acquired desires, intentions, and beliefs.

There is, however, a stronger way of understanding our third condition. On this reading, it's not enough that your belief that *C* is a reason. You have to see it, take it, or think of it as a reason. Here the notion of a reason, or some related notion, does have to enter into the contents of your thoughts. I'm not going to argue about which interpretation of the third condition really amounts to following a rule. As long as we see the difference, we can talk about them both. But I think the stronger understanding of following a rule gives you the notion of reasoning, and with it, the notion of an action.

So what does it mean to take something as a reason in the stronger sense? What is it like to have these thoughts about reasons running around in your head? And how often does this really happen? I think it happens all the time. For example, you're thinking in terms of reasons whenever you're thinking in terms of means and ends. It may be that instrumental reasons for *A*-ing are always about *A*-ing. I'm using a fairly inclusive notion of an instrumental reason, and there are at least three sorts of obvious examples. If you flip a switch in order to turn on a light, then flipping the switch causes the light to go on. Though flipping the switch doesn't cause your turning on the light, it's still a means to turning on the light. A second sort of obvious example is the preliminary step. If you turn on the light in order to put on your boots, or you put on your boots in order to go for a hike, here you do one thing and then you do another, but they're both part of the same plan. Finally, there's what we might call constitutive reasoning.[14] If you go for a hike in order to get some exercise, hiking is a way of getting exercise or it constitutes getting exercise. In all three cases where you *A* in order to *B*, your reasons for *A*-ing are about *A*-ing. A specification of the content of the beliefs or desires makes reference to the action type.

One of the most striking things about instrumental reasons is their diversity. What do all of these things have to do with each other? What else, if anything, goes on the list? What holds these notions together is the notion of a reason. To see *A* as a means to *B* is to see the connection between *A* and *B* as a reason for *A*-ing. To see *A* as a means rather than merely as a cause or a necessary or sufficient condition is to see it as something you have reason to do. I think this explains people's intuitions about intending and foreseeing. You might know that if you fire your gun you'll make a loud noise, but you don't take its making a loud noise as a reason to fire your gun. So firing your gun is not a means to making a noise. So while you may foresee making a noise, you don't intend to do it.[15]

If the notion of a means is itself a normative notion, if you understand it in terms of the notion of a reason, then for instrumental reasons, taking something as a reason in the stronger sense is simply thinking in terms of means and ends rather than causes and effects, preliminary steps, or constitution. Of course, your belief about means, or reasons, is grounded in your belief about causation or the rest. A complete theory of reasons or rationality would justify this move from is to ought. It would tell us why these things, but not various other things count as

[14] David Wiggins, 'Deliberation and Practical Reason', *Proceedings of the Aristotelian Society,* 76 (1975–6): 29–51.

[15] This is not to say that the distinction is morally relevant. The distinction between intending and foreseeing comes up in action theory when people ask if you can intentionally A without intending to A. There are intuitively plausible cases where you foresee A-ing without intending to A and in which it seems that you intentionally A. But there's a temptation to question the significance of these intuitions in the absence of an account of what the difference between intending and foreseeing comes to. Thinking of means in terms of reasons might provide such an account.

reasons. But here I'm only concerned with the distinction between something's being a reason, a fact I take for granted, and your seeing that thing as a reason.

When you *A* in order to *B*, you have to see the connection between *A* and *B*. Seeing the connection between *A* and *B* is a matter of having some thought about *A* and *B* under those descriptions. This thought may or may not involve normative notions like the notion of a reason. If you're thinking in terms of causes, conditions, or constitution, these are not normative notions. You might think that turning the knob on your stove will make a clicking noise, and you might think that turning the knob on your stove will result in turning on the burner. Both of these involve seeing the connection between turning the knob and something, but the connection seen is not a normative one. When you see a non-normative connection between turning the knob and making a noise, you may or may not think about turning the knob as something you do on purpose. You might think about it that way, but there's no guarantee that your belief that turning the knob will make a noise is all that much different from your belief that falling off a cliff will hurt a lot. The latter belief does not, either implicitly or explicitly, conceive of falling off the cliff as something you do on purpose.

Supposing for the sake of argument that intentionally turning the knob on your stove is a first-order propositional attitude, it doesn't follow that turning the knob on your stove is a propositional attitude. You can turn the knob by mistake and without even knowing what a knob or a stove is. So if the connection you see between turning the knob and making some noise is not a normative one, then seeing that connection may be a matter of having a first-order thought, and it might not involve any thoughts about your own mental states or events, even in my extended sense. But if you see the right kind of normative connection, if, for example, you see its turning on the burner as a reason to turn the knob, then you must, at least implicitly, conceive of your turning the knob as something you'd be doing for a reason, and so, something you'd be doing on purpose. So when seeing the connection is a matter of seeing something as a reason, your thoughts are not only normative, they're partly second-order. I take it that this kind of second-order normative thinking really does happen all the time. When you're planning your day, you really do, at least implicitly, think about that trip to the bank as something you'll be doing on purpose. And at least in normal cases, you know why you'll be doing it, that is, you know the reasons for which you'll be doing it.

But you can't always understand taking *X* as a reason to *A* in terms of seeing *A* as a means to *X*. Not all reasons for *A*-ing are instrumental reasons, even given the fairly inclusive notion that we're using. You can have reasons for trying to *A*, deciding to *A*, and concluding that *p* where nothing in the content of the reasons makes reference to the relevant propositional attitudes. Your reasons for deciding to go to the bank are your desire for money and your belief that you can get some at the bank. There's nothing in there about deciding. If you are just thinking about money and the bank, your thinking is first-order and

non-normative, and the intention, if there is one, is passively acquired. The intention may be rationally acquired, and the reasons that move you may be perfectly good reasons. You don't need higher-order normative thoughts in order to be reasonable. Rational causation may be enough for that. But you do need them to actively acquire intentions.

If your need for money and knowledge about banks are reasons for deciding, it seems that you could take them as reasons for deciding. And this mental state, seeing them as reasons for deciding rather than as reasons for believing or hoping is a higher-order, normative mental state. The intention is actively acquired or the decision actively made only when you see the normative connection between the reasons and the decision, so only when you're thinking in terms of reasons and decisions, so only in the presence of higher-order normative thoughts. Of course, seeing the connection between the reasons and the decision doesn't cause the decision. It just is making the decision. When you see that in the circumstances, these are good enough reasons to decide to *A*, that just is deciding to *A*. Again, this is not a claim about an entailment between the type, *seeing things as reasons to decide* and the type *deciding*. Of course it's possible to see things as reasons to decide without deciding. You might think there are, or even just leave open the possibility that there are other reasons not to decide. Or you might just be unreasonable.

So first, we have a claim about tokens, about what's what in the circumstances. In ordinary circumstances, seeing something as a sufficient reason to consider, wonder, decide, or conclude doesn't cause the considering or what have you. It just is the considering or what have you. This in no way commits you to any kind of necessary connection between the types. Second, we have a claim about what it is for the mind to be active. Basically, when raising the question just is seeing the reasons for raising the question, then raising the question is active. And the same goes for considering, deciding, and concluding. Since the claim that you have a reason to *A* is not a further reason to *A*, since your only reasons for deciding to *A* are the first-order propositional attitudes that you take to be reasons, we don't violate our rule about orders. Seeing something as a reason doesn't give you a further reason, but it does make the difference between rational causation and reasoning. If an action is always a step in the process of reasoning, then seeing something as a reason makes the difference between things that happen to you and things that you do.

In the case of mere rational causation, you're moved by something that is a reason, and this may be enough to make what you're doing reasonable. In the case of reasoning, you're moved by the fact that you see it as a reason. Of course, if you see it as a reason because it is a reason, you're moved by a reason in this case as well. So what does seeing it as a reason really add? One thing it adds, at least when things go well, is knowledge of what you're doing and why. This might not be sufficient for control over what you're doing, but it probably is necessary.

So on this view, it seems that reason alone is enough to move the mind. I'll let you decide for yourself whether this is a cost or a benefit. But if we grant this, it's no longer odd to think that, at least sometimes, belief formation is an intentional action. If you take your beliefs that p and that if p then q as reason to believe q, you might manage the action of concluding without the help of a desire. If the desire is first-order, its content is a logical truth, and it's hard to see how this desire could help or how its absence could hurt. If the desire is second-order, if it's about believing that q, then in the first instance, it justifies something second-order like wanting to believe that q. If you go for a walk just because you want to, that's fine. But if you believe that q because you want to believe that q, that's not so good. In any case, adding the second-order desire merely turns theoretical reasoning into practical reasoning ('reasoning about what to believe'). But then you'll need rules for practical reasoning, and desires are inputs into these rules, not the rules themselves. If adding the desire won't help, and concluding is something you do, then you have purely theoretical reasons for action.

What goes for concluding goes for deciding as well. Suppose you want to ϕ, and you believe that A-ing is constitutive of ϕ-ing or that A-ing is a preliminary step to ϕ-ing. Here you have instrumental reasons for A-ing. But your reasons for A-ing are your reasons for wanting to A, and they're your reasons for deciding to A. If your belief that these things are reasons for deciding can't move the mind, it's hard to see how adding some higher-order desire could help. If you desire to be rational or to think in accord with the principles of instrumental rationality, and you see deciding to ϕ as constitutive of one of these things, then this is just another instance of the very same rule. If you're sufficiently rational to respond to reasons in the second case, there's no reason to think the higher order desire is necessary. If you can do it in the second case, you can do it in the first. If you're not sufficiently rational to respond to reasons in the first case, there's no reason to think the higher-order desire will help. Since it's the same ability, if you can't do it in the first case, you can't do it in the second either.

So how often does this sort of higher order, normative thinking go on, the kind of thinking that makes decisions, judgments, and the rest active? Here it's not enough that you think about that trip to the bank as something you'd do on purpose. Here you need to think about deciding or judging, and you need to think about things as reasons. But when you're planning your day or trying to figure something out, aren't you just thinking about the world? In an ordinary, non-momentous case of planning, you might think about money and the bank, and you might even think about things that are in fact reasons, but do you really think about them as reasons and do you really think about deciding? I think you do. I think this higher-order normative thinking happens a great deal more often than it seems.

In case you're worried that I'm doing your introspecting for you and finding more there than you do, let me say a word about the kinds of thoughts I'm

talking about. To use a familiar example, when you're driving and talking to a friend, the focus of attention is on the conversation, not the road. But it's hard to believe that you'd make it home safely unless your beliefs about cars and curves and conditions played a role in guiding your driving behavior. This may not be enough to make these thoughts conscious, but it is enough to make them occurrent, and it's the latter kind of thought I'm talking about. If you look into your mind on the way to the bank, maybe all you'll find are thoughts about lunch. But if God looked into your mind, he would, no doubt, find a whole lot more.

Since I can't look directly into your mind, I need some evidence for the existence of these higher-order normative beliefs. We believe in the beliefs about the cars because we don't think that you could do what you do without them. And the same goes for planning your day. Planning isn't or at least isn't always a matter of free association. The possibilities you consider, the questions you raise, and the facts you take into account are all far more relevant to your project of planning than a random or undirected collection of thoughts would be. This can't just be a coincidence, and the idea that there are thoughts about planning somewhere in the background seems like a pretty good explanation. Though it might not happen as often as we would like, it really does happen all the time that things that are reasons get treated as reasons. It's hard to believe this would happen as often as it does unless people recognized them for what they are. So it looks as though there are thoughts about reasons lurking in the background along with the thoughts about planning.

So the phenomenology of planning might not count against my view, but surely young children and people without the relevant concepts couldn't have the relevant thoughts, even if the thoughts aren't conscious. I think that's right, and I think that's exactly as it should be. My eighteen-month-old is a very active boy (in much the way a volcano is active). On his way to grab the ball, he sees the block, which makes him change course, which brings the spatula into view, which he grabs and bangs on the floor while screaming with glee. I have no doubt that his path through the kitchen playground is the causal result of information about his immediate environment along with desires for things that he sees. What I seriously doubt is that this is all there is to the kind of activity characteristic of normal human action. He may well be moved by reasons. He grabs the spatula because he wants to play with it. But he won't be ready to act on reasons until he's ready to recognize them as such.[16]

[16] I read a half-baked version of this chapter at the University of Michigan in 1999. I'd like to thank the audience both for their patience and their comments. I'd like to thank the editors of this volume for encouraging me to give it another try, and I'd also like to thank them, and two referees for Oxford University Press, for helpful comments on a more recent version of it.

5

Reason, Voluntariness, and Moral Responsibility

Thomas Pink

1. ACTION AND RATIONALITY

Action is not something arbitrary or groundless. There are justifications or reasons why we should do some things and avoid doing others, or so we ordinarily suppose. In which case there is such a thing as a practical or action-directive reason. As adult humans we have the capacity both to recognize this directive reason in the specific justifications that constitute it, and to be guided through that recognition into performing those actions which are sensible and supported by reasons, and refraining from those actions that the available justifications oppose.

So our capacity for action, it is natural to think, is one part of our general rationality. To act is one way in which we can exercise our capacity to respond to reasons. This is something that we can do competently, when we act sensibly and in ways that are justified; or incompetently when, despite our sensitivity to reason and our general capacity to do what is sensible, we ignore reason and do what is foolish instead.

But action is only one mode by which, competently or incompetently, sensibly or foolishly, we exercise our rationality. For we can also arrive at beliefs, experience desires, and come to feel emotions. And all these psychological states or attitudes are similarly governed by reason, and can also be formed by us in response to reasons for and against. The evidence drives me to arrive at a particular conclusion about what happened, the tempting attractions of an offer bring on a desire to accept it which for a while almost grips me, the obvious danger of my situation overwhelms me with fear. These situations involve perfectly good justifications for belief, for desire, for fear. And in forming the belief, the desire, the fear, I am again exercising a capacity to respond to these justifications—to exercise rationality. But I am not performing any action.

When, for example, on being made a tempting offer I immediately feel a strong desire to accept it, my reason certainly is involved. I can be responding as a rational animal to a perfectly good rational justification—to features of the offer which do make accepting it, in some respects at least, very desirable. And I can be doing so in a process of deliberation or practical reasoning about whether to accept the offer. But feeling this strong desire to accept the offer is hardly going, just on that basis, to count as my own deliberate doing. The desire to accept is surely something which just comes over me—which my receptivity to these justifications just lands me with. Simply learning of the tempting details of the offer is enough to leave a desire to accept it arising irresistibly within me. That such a desire arises within me constitutes no intentional or deliberate doing on my part; nor would the failure of such a desire to arise constitute an omission of action.

Desire is a mode of exercising rationality. Desires can be sensible or foolish, and there can perfectly well be reasons for and against wanting things. But desire is nevertheless a mode of exercising rationality that is *passive*. By *passive* I mean here merely that the event of coming to have a desire is not an action, and correspondingly that the persisting state of desire that results is not a state whose beginning constitutes action. As passive, desires are things which happen to us without being directly our doing.

If action is one mode of exercising rationality, how does it stand out? What might distinguish action from these other modes of exercising rationality—from ways of exercising reason that are passive in comparison?

2. ETHICAL RATIONALISM

There is one field in which this question obviously matters—and that is in ethics. For morality seems centrally concerned with how we act. Moral standards seem to involve reasons for acting this way rather than that. If we were not capable of performing actions at all, but were merely passive observers of the passing scene, it is hard to imagine us being answerable to moral standards, or subject to moral appraisal and criticism.

Our moral responsibility, in particular, is for action. We hold most adult humans, ourselves included, responsible for how they act. And this responsibility is assumed in particular by one kind of moral criticism—blame. If what someone does is morally wrong, and so breaches moral standards, we may blame them for having done it. And the message of blame is that not only did they do wrong, but that they did wrong was their fault. Not only was what they did bad; but it was bad of them to do it. They were responsible.

Moral responsibility is for action. It is for how we act that we are responsible—not for what happens to us independently of our own doing. This is our natural intuition. But the intuition is often challenged. And the challenge

is based on a philosophical doctrine about action and, especially, its ethical significance. The doctrine is that, as far as our moral accountability for what we do is concerned, there is nothing to distinguish action and our responsibility for it from any other mode of exercising rationality. For if responsibility is understood properly, we are equally responsible, not just for what we deliberately do or refrain from doing, but for any rationally appraisable attitude. And so, recently, T. M. Scanlon has argued. Responsibility, he thinks, is not for actions and omissions alone, but for something more general. Responsibility is for how we exercise our capacity for rationality—something we do not just in our actions but more generally in forming psychological attitudes of belief, desire, emotion, and the like. What we are responsible for is how we exercise our reason: ' ''being responsible'' is mainly a matter of the appropriateness of demanding reasons.'[1]

So moral responsibility is something that we possess not just for our actions but also for our prior passive attitudes:

> For this reason, one can be responsible not only for one's actions but also for intentions, beliefs and other attitudes. That is, one can properly be asked to defend these attitudes according to the canons relevant to them, and one can be appraised in the light of these canons for the attitudes one holds. The 'sting' of finding oneself responsible for an attitude that shows one's thinking to be defective by certain standards will be different in each case, depending on our reasons for caring about the standards in question. But the basic idea of responsibility is the same. (Ibid. 22)

In fact, on Scanlon's view, it is primarily attitudes that are rationally appraisable, and it is primarily in attitude formation that we exercise our reason: 'Judgment-sensitive attitudes constitute the class of things for which reasons in the standard normative sense can sensibly be asked for or offered' (ibid. 21). Whereas actions occur only as expressions of prior attitudes, and it is only as such expressions that they count as rationally appraisable at all:

> Actions are the kind of things for which normative reasons can be given only insofar as they are intentional, that is, are the expression of judgment-sensitive attitudes . . . it is the connection with judgment-sensitive attitudes that makes events actions, and hence the kind of things for which reasons can sensibly be asked for and offered at all. (Ibid. 21)

The implication for moral responsibility is clear. According to Scanlon, what we are responsible for is not agency as such, but the rationally appraisable. Since our actions are rationally appraisable only as expressions of priorly rationally appraisable attitudes, Scanlon's view implies that we are responsible first of all for attitudes as things which occur within us prior to and independently of how we act; and we are responsible for our actions only as expressions of these prior and passive attitudes. Responsibility for non-actions comes first. And then responsibility for actions follows.

[1] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 22.

But I suspect that this is really the reverse of our ordinary view. For surely we ordinarily think that our moral responsibility is primarily for how we act—for what we ourselves do or fail to. It is what we ourselves do or refrain from doing and the consequences of this that are our responsibility—not what happens independently of our own doing. In which case any responsibility we might have for our passive attitudes is derivative, and based on the extent to which we have been in a position to use prior actions to influence some of those attitudes. To make our moral responsibility for what we ourselves do a secondary and derivative case of responsibility—this is to turn our ordinary view of moral responsibility upside down.

Ethical rationalism is the view that the ethics of action are entirely to be explained in terms of general reason. Moral standards are one kind of rational standard among others, and they govern action as one mode among others of exercising reason. To meet and conform to moral standards is no more than to exercise a general capacity for rationality—our capacity to respond to any standard of reasonableness. And to criticize someone morally is just to criticize them for some form of irrationality or unreasonableness—some failure in the exercise of reason. In so far as he equates moral responsibility with rational appraisability (' "being responsible" is mainly a matter of the appropriateness of demanding reasons', ibid. 22) Scanlon is a clear ethical rationalist.

3. BLAME

But this equation of moral responsibility with rational appraisability must be a mistake, at least as an interpretation of our ordinary ethical thinking. For, as we ordinarily understand it, moral responsibility clearly is a far more demanding notion than mere rational appraisability, and in a way which presupposes action. What shows this is the nature of blame—the criticism we make of people when we hold them responsible for having done wrong.

When we blame people for wrongdoing, we are not merely alleging some fault or deficiency in their response to reasons. And this is because of something that has already been mentioned, and that is quite central to the content of blame. Blame does not just report a deficiency in the person blamed. It further states that this deficiency was the person's fault—that they were 'to blame' for it. The attribution of something not only as a fault, but as someone's fault, the fault of the person blamed, is essential to anything recognizable as genuine blame.

And this fact about blame is key to what distinguishes moral responsibility from any mere rational appraisability, and ties moral responsibility to being specifically for how we act. Blame, which asserts one's moral responsibility for what one is being blamed for, not only detects a fault. It also condemns it as one's own fault—and one's own fault as one's own doing.

For suppose someone is subject to ordinary rational criticism. Suppose, for example, that they have committed some error of reasoning. It is always a *further* question whether that they made this error was their fault. Are they responsible and to blame for the fact that they made it? Or did they make the mistake through no fault of their own? They were certainly being foolish or less than sensible; it is, after all, their reasoning which was bad. But we can still ask whether it was through their own fault that they reasoned incorrectly. The question of one's responsibility for one's attitudes remains open, even when one's rational appraisability for those attitudes is admitted. In which case Scanlon must be wrong. The kind of responsibility assumed in blame does not reduce to the appropriateness of rational appraisal.

R. M. Adams makes the same mistake as Scanlon—that of confusing blame with more general rational or ethical criticism. In his 'Involuntary Sins' Adams has suggested that people can be blamed for attitudes, such as selfish motivations, that are not of their own doing.[2] Now certainly people can be criticized simply for being selfish. And Adams supposes that such criticism always amounts to blame:

> Perhaps for some people the word 'blame' has connotations that it does not have for me. To me, it seems strange to say that I do not blame someone though I think poorly of him, believing that his motives are thoroughly selfish. Intuitively speaking, I should have said that thinking poorly of a person in this way is a form of unspoken blame.[3]

But contrary to what Adams claims, to be criticized as selfish does not itself amount to being blamed. For again the selfishness of someone's motivation does not of itself settle the question which is raised in blame—namely their responsibility for the motivation they possess. Their selfishness is one thing. It involves a failure to respond to the reasons for motivation and for action provided by other people's interests. It is still a further question whether their possession of such a character is their fault. For there is no inconsistency at all in criticizing someone as having a selfish character, while wondering or doubting whether their possession of this character really is their fault. Just as someone can be criticized as poor at reasoning without this being supposed to be their fault, so they can be criticized as selfish without this being supposed to be their fault.

Failing to respond in one's beliefs or motivations to reasons is one thing. Being responsible for that failure as occurring through one's fault is another. What settles whether someone is so responsible? How, for example, to establish that it was through the bad reasoner's own fault that their reasoning was faulty? It is obvious enough how. We would raise questions about their action and omission and about how this might have affected their responsiveness to reason in this case—questions such as the following: what if they had taken greater pains at the time, such as by attending more carefully or taking longer to reflect; or had

[2] R. M. Adams, 'Involuntary Sins', *Philosophical Review*, 94 (1985): 1–35. [3] Ibid. 21.

prepared themselves better beforehand, such as by working harder at practicing this form of reasoning? Did the error arise from their failure to do any of these things? If on the other hand they would make the error whatever care they took, if they are simply not good enough at this kind of reasoning, then their making the error, though it certainly lays them open to rational criticism, is clearly not their fault. And so generally: we are responsible for a faulty response to reason only if this faulty response is brought about as or through some action or omission of our own. We are responsible for it only if it arises as our doing.

Responsibility of the sort invoked in blame, the idea of something's being one's fault, really is tied to agency. We can always sensibly ask whether something already admitted as mistaken or bad, such as a reasoning error or a selfish disposition, really has arisen through the person's own fault. And the answer to the question then depends, at least in part, on doing. We are responsible for the bad occurrence only if it arises as or out of our own action or omission.

4. SELF-DETERMINATION

It seems then that moral responsibility is for how we act, and for other things as consequences of how we act. But if moral responsibility really is for agency, what might explain this tie? Not the fact that action is a mode of exercising rationality, for that can equally be true of non-actions. There must be something else about action—something which distinguishes it from non-action.

Once more consider blame. When we blame someone, we put the fault we are blaming them for down to them, as their fault. This imputation of something to the person blamed as being their fault is, I have said, absent from mere rational criticism, but is essential to blame. And it is key to understanding the link between moral responsibility and action. For it uncovers a link between moral responsibility and a power—a power of *self-determination*.

What the person is blamed for is put down to them as their fault—and their fault as a mere failure of rationality need not be. But if they are to blame this means that they must have been in a position in some way to determine for themselves whether it occurred—and determine in a way that goes beyond any mere exercise of their capacity for rationality. Why else would the thing be their fault as an ordinary exercise of rationality need not be? So at issue then is some power that they possessed over what they are being blamed for—a power distinctively to determine its occurrence.

If blame is for how we act, then that will be because this power to determine things for ourselves applies specifically to how we act, but not to the exercise of our rationality in general. If we are directly morally responsible for our actions and our failures to act, but not, say, for non-actions such as, say, mere desires, that is because the power of self-determination which moral responsibility presupposes is exercised in and through how we act, but is not exercised simply in coming to

have a desire. It is this power and its link with action which explains our special and moral responsibility for how we act. If we exercise this power, or if, despite possessing it, we fail to exercise it, then we ourselves can be responsible and can be truly to blame. What happened as a result can truly have been our fault.

At the heart of our ethical thinking there lies an idea of power. Our moral responsibility for how we act presupposes a kind of power to determine for ourselves how we act. The bearer of the power is a person; and action is both the object over which this power is immediately exercised and, in so far as we determine other things through how we act, a medium for the exercise of this power over those other things. It is the existence of this power which explains, if anything does, why it should in particular be for our actions and omissions of action that we are responsible.

Our natural conception of this power is, of course, freedom. Freedom is the power that we report when, contrasting action to the desires that come over us, or the beliefs about the world that memory and our senses impose on us, we claim that within the general limits set by our intelligence, strength and resources, how we act is up to us. Freedom or the up-to-us-ness of our action is the power which we think we possess to determine for ourselves which actions we perform. It constitutes a control over how we act, leaving alternatives by way of action available to us, so that we determine which of these alternatives we do. If we blame agents—ourselves and others—for actions that they perform, that is because how the agent acted was the agent's fault; and it was the agent's fault because they were in control.

Notice that freedom, if it does exist, must be very different from the mere capacity for rationality. That must be so, otherwise freedom could not explain why it should be for one mode of exercising rationality in particular—the performance of intentional actions—that we are peculiarly responsible. But there is also a deeper reason why the power to determine our actions for ourselves must be distinct from rationality.

Self-determination constitutes a form of power in that it constitutes a capacity to determine what happens. For a capacity to determine what happens just is what power amounts to: *A* has a power over *B* if *A* has a capacity to determine what happens to *B*. But rationality is not a case of power in this sense at all. It is not itself a capacity to determine. In fact rationality is a quite different kind of capacity. It is a second-level capacity—a capacity exercised in relation to other capacities, such as for desire, belief, and action. Rationality is a capacity to exercise those capacities in a certain way—a way that is responsive to reasons or justifications. And these capacities that are to be exercised rationally need not themselves involve the possession and exercise of power over anything. Far from being capacities to determine, they may rather be capacities to be determined. In which case the exercise of rationality may be a mode, not of exercising power, but of being subject to it; and rationality will be exercised in a way that excludes control rather than permits it.

Take everyday belief for example. When I believe, as I do, that I am sitting in my study and am surrounded by tables, books, and chairs; that outside my study, and extending far beyond what I can presently hear or see, is a whole city with millions of people in it—all this is a perfectly good exercise of my capacity for rationality. These beliefs that I form are a fully reasonable response on my part both to the evidence of my present experience and to what I remember of the past. But I certainly do not have any control over whether I form these beliefs. It is not up to me whether I believe that there are chairs in my room and that there are millions of people outside.

What leaves what I believe so clearly outside my control? My own capacity for reason is a crucial part of the story. It is my very rationality, given the nature and function of belief, the capacity that is being exercised so rationally, which imposes these beliefs upon me. My rationality imposes these beliefs on me as so obviously true that it simply is not within my control to think otherwise. In the case of these beliefs about my surroundings, far from freedom being the same as reason, freedom—the freedom to believe otherwise than as I actually do—is something which reason helps prevent. For in relation to beliefs such as these, reason governs not a power of self-determination, but a capacity to be determined—by the facts or the reality that the beliefs are about. Rationality ensures that this capacity functions properly so that, far from what I believe being left up to me, my beliefs faithfully track the evidence. Thanks to my rationality, my beliefs are determined, not by me, but by the very reality they serve to represent.

So exercising rationality has nothing directly to do with exercising power—with determining as opposed to being determined—unless, of course, the particular justifications to which one is responsive are justifications for exercising a power. But if they are, then the power, such as the power of freedom, has to be supplied. It is a further and distinct element beyond the capacity for rationality.

Modern ethical rationalism, we have seen, ignores the possibility of any agency-specific power of self-determination. According to the ethical rationalist, moral standards are just an instance of standards of general rationality, and moral criticism is just another instance of rational criticism. There is no kind of ethical criticism, such as blame, nor any kind of ethical standard, such as moral obligation, that might presuppose some special power to determine things for oneself. But given that we clearly *do* have a practice of blame that presupposes a power of self-determination going beyond any mere capacity for rationality, why has ethical rationalism gained so many supporters? One important reason is the currency of a theory of action that leaves the reality of self-determination in any ethically significant form very problematic. The exercise of rationality in action looks as though it is left a simple expression of its exercise in passive form. Action is left no more than an expression of the passive.

5. PRACTICAL REASON, MOTIVATION AND VOLUNTARINESS

Consider how we as rational animals respond to practical reason and apply it in action. This process involves two stages: *will* or *motivation*, and then *voluntariness*.

Motivation involves the formation of various pro attitudes towards actions and outcomes. These motivations are content-bearing psychological attitudes which leave us favouring the actions and outcomes specified by their contents—the objects of those attitudes. Motivations include not just desires and passions, but intentions too. Intentions are attitudes which constitute our final overall motivation to act, thus determining what we finally do, and which we form in taking decisions to do this rather than that. For example, we first form an intention to visit a shop across the road. Then, believing that now crossing the road here is the best means to this end, we form an intention so to act—to cross here and now. And so we are left fully motivated to cross.

Voluntariness then involves actually doing things on the basis of the pro attitudes towards doing them that we have formed. So when I talk of actions that are *voluntary* I shall mean actions that are performed in just this way, on the basis of some prior motivating pro attitude towards performing the action, such as a prior decision or intention to perform it. Thus once I form my intention to cross now, then—unless some incapacity or obstacle prevents it—that motivation will, through its effects on the motion of my limbs, lead me actually to cross. In which case I can count as crossing voluntarily. Notice that it is far from obvious that all actions need be voluntary in this sense. Voluntariness is certainly one form taken by intentional action. But it may not be the only one; though, as we shall see, it is a central assumption of much modern English-language action theory that action and voluntariness come to the same thing.

Deliberation or reasoning about how to act—practical deliberation—is principally and centrally about which voluntary actions to perform, such as about whether now to cross the road here, and the like. In deliberating we consider the various features which these possible voluntary actions have. We consider the actions both as possible ends in themselves, things possibly worth doing for their own sake, and as possible means to attaining further ends. Certain features of the voluntary—such as, say, the fact crossing the road here would get me to the shop on the other side—then generate reasons or justifications for performing this voluntary action rather than another.

Take the fact that crossing the road would get me to the shop on the other side. Suppose this feature of the voluntary action is a justification for performing it—for crossing the road. If this justification is to move me to act, it cannot

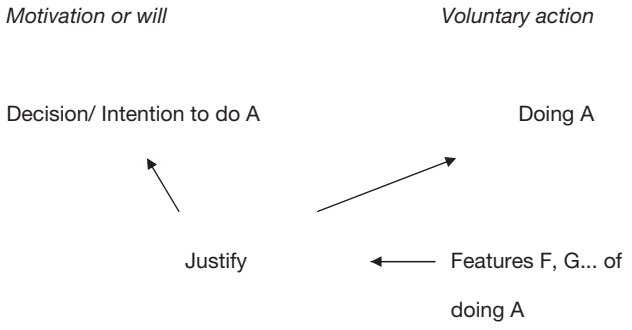


Fig. 5.1. Motivation and action

apply to the voluntary action of crossing the road alone. It must apply also to the will or motivation on which the performance of that voluntary action depends. The justification must also support deciding to cross the road.

If practical justifications could not address our will or motivation just as they address the voluntary—if they could not provide the same support for being motivated to do *A* which they provide for doing *A*—they would bypass the motivation which determines how, at the level of the voluntary, we act. Though one had every justification for doing *A*, one would not have the same justification for deciding or becoming motivated to do *A*. In which case, no matter that one was perfectly rational, and fully capable of responding to justifications, these will by-passing justifications could not move one into performing the actions which they justify. One would note the justification for action; but no matter that one was perfectly rational, and fully capable of acting as justified, one would simply ignore the justification in one's decision-making. But that would be absurd. For it cannot be in the nature of practical justifications simply to pass rational agents by in this way. It cannot be in the nature of practical justifications to lack the force to move us to do what they justify doing.

The desirable features of a given voluntary action, then, only count as justifying its performance insofar as there is also the possibility of their providing the same justification for being motivated—deciding or intending—to perform it. Justifications for performing a voluntary action must be able to constitute equal justifications for the intention so to act. Only so can reason ever direct and move us into performing this voluntary action rather than that.

Our response to practical justifications occurs then at two points. The first and primary point is in our motivations—the point at which we decide or intend to perform this voluntary action rather than that. It is at the point of the will that, on the basis of our deliberations, we determine which voluntary action

we will perform. That is the whole point of deciding and forming intentions about which voluntary actions to perform—to settle which such actions we shall perform. Then through the action-determining effects of our intentions we end up actually performing the voluntary actions for which we have justification. If we really do intend to perform an action now, then through the effects of that intention we will so act—unless some obstacle or incapacity prevents. Our response to justifications in voluntary action is thus an expression and effect of a prior response to those same justifications through our will.

Not only do our motivations constitute our immediate response to practical reason, and thereby serve to determine which voluntary actions we perform. Those motivations also provide our voluntary action with something both fundamental to its character as action, and essential to its rationality. And that is its goal-directedness, which in the case of voluntary action comes from the contents of the attitudes that motivate how we voluntarily act. For our motivations determine the object at which our voluntary action is directed—the goal that we are aiming at in its performance. As I now move across, what makes it true that I am intentionally crossing the road now in order to get to that shop on the other side? That in so moving I am acting on the basis of an intention to cross now, which is in turned based on an intention to get to that shop. The contents of my intentions, my pro attitudes towards what I am voluntarily doing, supply my voluntary action with the goals at which it is being directed. And that goal-direction then helps determine how rational the action is. The voluntary action will be being performed rationally only if its goals are sufficiently desirable and sufficiently likely to be attained by the action—thereby justifying both the voluntary action and the intention to perform it.

This is why Scanlon is importantly right on one point at least, that the rationality of voluntary action is a function of the rationality of prior motivating attitudes: voluntary actions

> are the kind of things for which normative reasons can be given only insofar as they are intentional, that is, are the expression of judgment-sensitive attitudes . . . it is the connection with judgment-sensitive attitudes that makes [such] events actions, and hence the kind of things for which reasons can sensibly be asked for and offered at all.[4]

Justifications for voluntary action are responded to by us as justifications also for prior attitudes; and the goal-direction and rationality of our voluntary actions depends on the content and rationality of prior motivating attitudes.

Nothing in this picture yet forces us to espouse ethical rationalism. Ethical rationalism only becomes attractive, even unavoidable, if we add to this model of action rationality a certain model of the nature of action itself.

[4] Scanlon, *What We Owe*, 21.

6. THE VOLUNTARINESS-BASED MODEL OF ACTION

This is a model of action which goes back to Hobbes. It is *voluntariness-based*, identifying intentional action exclusively with intentional action that is voluntary. Intentionally to perform an action *A* in order to attain some end is always to do *A* on the basis of, and through the effects of, a prior will or motivating pro attitude towards doing *A*.

On Hobbes's version of the theory, and on most other forms of it,[5] that motivating pro attitude is itself passive. The boundary between motivation and voluntariness is the boundary between non-agency and agency—between passive and active. The formation of a motivating pro attitude towards doing *A* is not itself a further case of intentional action; and this applies not just to intuitively passive attitudes such as desire, but to intention also.

And this is because the formation of a particular motivating pro attitude is not something voluntary. It is not itself motivated by prior pro attitudes towards doing it as is the voluntary action it explains. And this Hobbes himself pointed out, frequently and with characteristic vigor. As he famously put it, using the seventeenth-century term *willing* for our modern *deciding*: 'I acknowledge this liberty, that I can do if I will, but to say, I can will if I will, I take to be an absurd speech.'[6] So not only is coming to feel a desire not an action, according to Hobbes; but nor is the taking of a particular decision or the formation of a particular intention.

And Hobbes was surely right on one point at least. Motivations such as intentions and decisions to act are not voluntary. They are not directly subject to the will. Intentions cannot be formed or decisions be taken just on the basis of prior decisions or desires so to decide. For example, I cannot decide that in precisely five minutes' time I shall then take a decision to raise my hand—and sensibly expect that in five minutes, at the appointed time, I shall take the decision decided upon, and take it voluntarily, just on the basis of my earlier decision that I shall take it.

And this is connected with another feature of decisions which also distinguishes them from what we really can do voluntarily—from what we really can do on the basis of a prior decision or desire to do it. Just as decisions are not directly

[5] Davidson follows Hobbes, denying that intention-formation is itself an intentional action; and making this denial on the grounds that intention-formation is not voluntary as genuine action is: see his 'Intending', in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), 90, and pp. 213–14 of his replies in B. Vermazen and M. Hintikka (eds.), *Essays on Davidson* (Oxford: Oxford University Press, 1985). In n. 8 below I critically note some recent attempts, unsuccessful in my view, somehow to reconcile the intuitively active nature of decision-making with the voluntariness-based model.

[6] Hobbes, *Of Liberty and Necessity,* in *British Moralists*, ed. D. D. Raphael (Indianapolis, IN: Hackett, 1991), i. 61–2.

subject to the will, so they are not directly subject to command.[7] I cannot sensibly command you to take a particular decision, such as a decision to raise your hand, and then expect you to take the decision commanded exactly as commanded and simply in order to obey my command. Suppose, for example, I commanded you thus: 'In five minutes' time take a decision to raise your hand tomorrow—and then, after a further minute, abandon that decision, and instead decide not to raise your hand tomorrow. Then, after yet a further minute, abandon that decision too.' You would surely react to my command with some bewilderment. You would be quite incapable of carrying it out. Decisions are not things that can be taken simply in order to obey commands that they be taken.

The fact that decisions cannot sensibly be commanded is obviously connected with the fact that decisions cannot be taken voluntarily. For if decisions could be taken voluntarily, on the basis of a prior desire or decision to take them, then you could perfectly well take decisions just in order to obey my decision commands. You need only decide to take whatever decisions I commanded you to take, and then obeying my decision commands would be easy. Once I commanded you to take a particular decision, such as a decision to raise your hand, you would simply take that decision voluntarily, on the basis of a decision to take it, and as a means to fulfilling my command.

It is very clear why decisions cannot sensibly be commanded. To get you to take a particular decision, to decide on a particular action, I cannot just command you to take the decision in question. If I want to get you to decide to raise your hand, I cannot just command you 'Decide to raise your hand!' To get you to decide to raise your hand, I have somehow to convince you that acting as decided, raising your hand, would be a good idea. I have to give you some reason to raise your hand. I have somehow to show or make it clear to you that raising your hand would have benefits.

One way to do that, of course, is to *make* it true, and obviously true, that raising your hand would have benefits. For example, I could offer you a reward for raising your hand. That could get you to decide to raise your hand. Or if I have the necessary authority to do so, I could simply issue a command. Not a command to decide to raise your hand, but a command that would give you a reason to act as decided and actually raise your hand. I could simply command you to raise your hand. Given this command, there would be one possible benefit to your raising your hand, one reason why you should raise it—namely, that in doing so you will manage to obey my authority. And by giving you this reason to raise your hand, I could again get you to decide to raise it.

If anything moves one to take a particular decision then, to decide on a particular action, it is not any command to take that decision, or any prior desire or decision on one's own part to take it, but some reason for acting as

[7] As also noted in O'Shaughnessy, *The Will* (Cambridge: Cambridge University Press, 1980), ii. 300.

decided—something good or desirable about the action decided on. As with desires, so with intentions and the decisions that form them: the attitudes are responsive, not to the will or to their own desirability, but to the desirability of their objects. If anything moves one to decide to raise one's hand, it is not any command so to decide, or any decision on one's own part so to decide, but something quite different—something good or desirable, not about taking the decision itself, but about performing the action to be decided upon, about actually raising one's hand. So Hobbes was quite right about the non-voluntariness of decisions. We cannot take a particular decision to act at will, just because we have decided to take it.[8]

According to the Hobbesean voluntariness-based model of action, then, the motivations that move us to act are all passive. The formation of a particular pro attitude is not something we intentionally do; and so action only occurs at the point of the uncontroversially voluntary, and as an effect and expression of the passive.

It is this theory of action which takes self-determination out of ethics, and leads to ethical rationalism with its reduction of moral responsibility to no more than a general rational appraisability. And the theory of action threatens to exclude

[8] We should not be distracted from appreciating this by the fact that there is something else connected with decision-making that we can do voluntarily. This is not taking a particular decision to act, but something which is very easily confused with it but which is, nevertheless, importantly different—namely making up our mind one way or the other what to do.

I can perfectly well decide today that in five minutes' time I will make up my mind one way or the other about whether to raise my hand—and then, in five minutes' time, make my mind up on the basis of that earlier decision. Making one's mind up one way or the other, then, is something that can be done voluntarily, on the basis of a prior decision to make one's mind up. But that does not show that the taking of a particular decision can be voluntary too. Making up my mind, after all, is a process. And it initially involves deliberating or at least seriously considering the options—something which is done preparatory to making any particular decision. It is this which can be done voluntarily. But which particular decision I then arrive at is not a voluntary matter. If when I make my mind up I decide to raise my hand rather than lower it, this is not something I can have done on the basis of some earlier decision to arrive at that particular decision rather than its opposite. I can decide in advance *that* I shall make my mind up; but I cannot effectively decide in advance *how* I will make my mind up. Deciding to do this rather than that is something entirely non-voluntary, just as Hobbes supposed. Deciding to do this rather than that is something I do in response to the options as I see them at the time—in response to their potential benefits, real or apparent—and not on the basis of some earlier decision to decide this particular way rather than the other.

Some writers attempt to defend the existence of an intentional agency of decision and intention-formation within the terms of the voluntariness-based model. They do this by adverting to the voluntariness of making one's mind up—see Randolph Clarke in *Libertarian Accounts of Free Will* (Oxford: Oxford University Press, 2002), 26; Robert Kane in *The Significance of Free Will* (Oxford: Oxford University Press), 138–9; and Alfred Mele in 'Agency and Mental Action', *Philosophical Perspectives,* 11 (1997): 231–49, 'Deciding to Act', *Philosophical Studies* (2000), and in *Motivation and Agency* (Oxford: Oxford University Press, 2003), 205.

But the voluntariness of making one's mind up only establishes that making one's mind up can be something we may do intentionally; not that taking a specific decision to do *A* is something we may do intentionally. To establish that in voluntariness-based terms, we would of course need to establish something very different—the voluntariness of taking a specific decision.

self-determination from ethics quite decisively, by making any substantial, agency-specific self-determination impossible.

It threatens to do this in one very clear way—by pushing the self-determining agent out of the picture. For according to the theory, it is through prior motivations, through passive attitudes that move us to act, that action occurs. By its very nature action occurs as the expression and effect of a lot of events and states not themselves of the agent's own intentional doing. It is these events and states that give rise to the action, and explain why the agent does what he does by giving what he does its goal-direction and purpose. In which case it looks as though it might really be these events and states, and not so much the agent, that determine action. The action-determining agent threatens to be pushed aside by the passive attitudes that define the very nature of action—and define it precisely as those prior factors that determine how the agent acts, and which explain why the agent performs the actions that he does.

There are two obvious but very different ways of defending the reality of self-determination within the voluntariness-based model's terms. First, one might accept that, in the context of the model, the action-determining role of the agent's motivations is indeed a threat to his own action-determining role. To make room for the agent as determinant of his own action one must limit the causal influence of his motivations on what he does at the point of the voluntary. The agent's motivations must at most influence, not determine, how the agent acts, leaving the final outcome for the agent to determine.

But this has an unpleasant consequence. The agent's power to determine what he does, on this picture, is being exercised apart from the very motivations by which he responds to and applies practical justifications. So the agent's power of self-determination, far from being exercised through and over his capacity for rationality, actually functions apart from it. But surely that is not what we ordinarily suppose. When I exercise control over what I voluntarily do, I do not do so apart from the capacity to respond to justifications that I exercise in my motivation and will. Rather I control my voluntary action *through* the will—through my decision-making. And that is what is at stake in the common thought (fiercely denied by Hobbes and many of his successors in the English-language tradition) that freedom of action depends on freedom of will. Or as we put the matter in everyday life, in what would ordinarily seem an unexceptionable truism—it's up to me what I do, because I can decide what I shall do, and it's up to me what actions I decide to perform. My freedom is exercised in and through my decision-making, not apart from it. If it is my own decisions that settle what I end up doing, if they do not merely partially influence the outcome but outright determine it, then so far from threatening my action control, this actually counts towards it.

An alternative strategy is to deny that the action-determining role of an agent's passive motivations is any threat to self-determination. The agent is to be identified with his motivations even supposing these to be passive—so that

the causal power of his passive motivations to determine how he acts is held to constitute *his* power to determine how he acts. And thus we arrive at the classical English-language compatibilist account of freedom—as a power of one's will or motivations to cause one successfully to act as motivated. I have control over whether I raise my arm or not just in case if I decided to move it I would do so, and if I decided not to move it I would do so.

The difficulty of this approach is that the agent seems very clearly distinct from any set of passive motivations that are not his doing. So why should we identify the agent with these motivations? Of course, in everyday life we do identify the agent with the operation of his will. But that is only because we ordinarily also think that, far from being a passive antecedent of action, the will is itself a locus of action, so that agents can and do directly control what they specifically decide, and therefore are fully responsible for their decisions and intentions. But, as Hobbes himself insisted, the classical compatibilist view does not allow for such a freedom of the will—unless, contrary to Hobbes, the will were itself voluntary, and we did have the capacity to take decisions at will, on the basis of deciding to take those specific decisions. But we do not.

But even if we do seek to understand self-determination as a simple power to act as one passively wills—as mere voluntariness—there is a further problem with seeing self-determination so understood as of any distinctive ethical significance, and so as the basis of a distinctive moral responsibility for how we act. It is not just that voluntariness is an alarmingly thin and insubstantial conception of self-determination, though that is certainly true. As we have noted, the agent does not seem at all the same as some set of motivations within him that are not his doing. The case against basing moral responsibility on voluntariness goes deeper. We cannot hold an agent morally responsible for the voluntary actions which his motivations determine without also holding him morally responsible for those non-voluntary motivations as well. That, at least, is the implication of a very natural intuition about what moral responsibility is for.

The intuition is this. If we are morally responsible at all, we are surely morally responsible for one thing in particular. We must be responsible for whether we respond or not to morality and to the justifications for action which morality provides. But, as we have seen, the point at which we initially respond to or disregard moral standards, and indeed practical justifications generally, must be, not voluntary action, but non-voluntary motivation. It is in and through our motivations that we conform to morality or disregard it.

Thus suppose someone's need justifies my relieving it through giving them help. I respond to that need and to the moral justification which it provides not simply by giving help, but by deciding to relieve the need. Their need is something I must genuinely intend to relieve; and it is on the basis of that intention that I must be providing the help. If my help is not based on such an intention, but is instead provided by me purely by accident, or only out of a decision to gain political kudos through being seen to help—then my help

will not count as a genuine response to the person's need and to what that need morally justifies. The agent's response to or disregard of morality and its demands begins in his motivations—in what he decides and intends to do—and occurs in his voluntary actions only as an expression of and through the effects of his decisions and intentions, effects that will follow provided no obstacle or incapacity prevents. Why then hold the agent morally responsible for these voluntary effects alone, and not for the decisions as well? If fundamentally it is through decision and intention that we respond to morality or disregard it, it must be for what decisions we take and intentions we form that we are morally responsible—if we are morally responsible at all.[9]

No wonder then that Hume insisted that the scope of blame and moral responsibility is not restricted to the voluntary and to outcomes of the voluntary, and that our moral responsibility begins with non-voluntary antecedents of the voluntary:

> Philosophers, or rather divines under that disguise, treating all morals as on a like footing with civil laws, guarded by sanctions of reward and punishment, were necessarily led to render this circumstance, of voluntary or involuntary, the foundation of their whole theory . . . but this, in mean time, must be allowed, that *sentiments* are every day experienced of blame and praise, which have objects beyond the dominion of the will or choice, and of which it behoves us, if not as moralists, as speculative philosophers at least, to give some satisfactory theory and explication.[10]

And about blame Hume is plainly right. We do blame people not just for the voluntary actions which their selfishness motivates, but for being selfish—for lacking any willingness and intention to respect and further the interests of others as well as their own. And we continue to blame people for their selfishness even when their selfishness has ceased to produce voluntary actions that harm others. We tell selfish Jack who is neglectful of his mother that it is very bad and wrong of him to be so selfish. We treat Jack's selfishness not only as something bad in itself, but as something that it is bad of him to be: we impute his selfishness to him as his responsibility. And we will continue to reprove him in such terms even once he begins to give his mother help—if it is clear that his only motive in so doing is still selfish, an intention to avoid being cut out of her will. What we hold Jack morally responsible for is whether he intends his mother's good as well as his own. And that is because we hold people

[9] In my 'Moral Obligation', in Anthony O'Hear (ed.), *Modern Moral Philosophy* (Cambridge: Cambridge University Press, 2004), and in my 'Reason and Normativity', *Journal of Moral Philosophy* (2007), I argue that moral obligation, the moral standard that we are morally responsible for keeping, consists in a distinctive kind of justificatory force—a force that by its very nature must lie not just on our voluntary actions, but directly on non-voluntary motivations of the will as well. In which case a direct moral responsibility for decisions and intentions follows not merely from general intuitions about responsibility, but also from the nature of moral obligation itself.

[10] Hume, *Enquiry Concerning the Principles of Morals*, ed. P. Nidditch (Oxford: Oxford University Press, 1975), appendix 4, 'Of some verbal disputes', p. 322.

morally responsible for whether or not they actually do respond to morality and its demands, and not simply for acting at the point of the voluntary as if they did.

Suppose then that moral responsibility cannot be exercised in voluntary action alone; that we must also be responsible for the decisions and intentions on the basis of which we so act. The question then moves to the precise basis on which this responsibility for our intentions is had. And here we have a choice. Either moral responsibility remains tied to agency—and to some power of self-determination exercised in and through agency. But then this power of self-determination cannot be voluntariness, since it must be exercised as much in and through the formation of non-voluntary intentions as in any voluntary actions which those motivations explain. In which case the power must be freedom in a form detached from voluntariness; and the formation of specific intentions must itself count as an intentional action.[11] Or we simply abandon the tie of moral responsibility to agency. In which case responsibility ends up being extended, as Scanlon wants to extend it, from actions and omissions to our reason-responsive attitudes generally. We can be morally responsible not only for how we act, but for our passive attitudes as well.

Theories which seek to tie moral responsibility to action via the notion of voluntariness are, in my view, profoundly unstable.[12] The notion which they appeal to is too weak to look convincing as a genuine form of self-determination. And it is anyway too restricted in scope to include all that we seem to be morally responsible for—a responsibility which extends to our decisions to act as well as the actions decided upon. Either we return to some other power, such as freedom, as the true explanation of the responsibility–action tie; or we abandon that tie altogether, and detach moral responsibility entirely from any dependence on agency-based self-determination. The serious debate, then, is between the protagonists of freedom as a condition of moral responsibility—who still wish to tie moral responsibility to agency-based self-determination of a substantial kind; and those who, like Scanlon, wish to equate the exercise of responsibility with the mere capacity for reason-responsive attitude formation.

[11] Freedom is a power that leaves it up to us, or within our control, how we act. Despite the marked tendency of many philosophers to see freedom so specified as consisting in or implying some kind of voluntariness, note that nothing in this idea of freedom obviously involves voluntariness, or the subjection of what we control to our will. Indeed it cannot do so if it is true both that we control what we decide to do, and at the same time that decisions are non-voluntary and not themselves subject to the will. If there is such a thing as freedom of will, freedom must be a power that extends over the non-voluntary. I discuss the nature of freedom itself and its distance from voluntariness, in my *The Ethics of Action: Self-Determination* (Oxford University Press, forthcoming); in particular I show there how the supposed linkage of freedom to voluntariness has arisen as a deep and powerful philosophical illusion.

[12] The project of basing our moral responsibility for action on voluntariness rather than on freedom has been a central theme in the work on moral responsibility of Harry Frankfurt and his supporters. I discuss Frankfurt's views in more detail in my *The Ethics of Action: Self-Determination.*

7. SELF-DETERMINATION AND THE WILL

As our consideration of blame has shown, at the heart of our ordinary thinking about action and its morality is the intuition that action really is ethically distinctive. To perform an action, actually to do something, is morally speaking quite different from simply being passively motivated so to act. And what makes action so very different is self-determination. When we act we are not only doing something, as opposed to having something happen to or within us. We are also, in general, exercising a power to determine for ourselves what we do. We can determine for ourselves what happens in our lives; and this we do in and through our very intentional agency itself—in and through what we deliberately and intentionally do or refrain from doing. And central to the exercise of self-determination is our decision-making capacity—our capacity to make immediate motivational response to practical justifications. Our free agency begins at the will, and is exercised in and through what we ourselves decide and intend.

But what makes self-determination possible? How can action have an independence from prior happenings not of our own doing—enough independence to be something genuinely determined by us, as opposed to being a mere expression of those prior happenings? And how can free action occur not just in voluntariness, but in our decisions too—in non-voluntary motivations of the will?

To answer these questions involves giving a very different account of what action is. And the required theory of action is to be found in the theory of rationality—in an account of how it is that reason is exercised by us in action.

Of course the self-determination which we exercise in action must, on my view, involve more than just the exercise of reason. This has to be true, as we have seen, both because self-determination involves power as rationality need not, and since otherwise self-determination and the moral responsibility which it bases would not be specific to how we act, but apply to any mode of exercising rationality. Nevertheless, it is still the theory of rationality which is key to uncovering the nature of the action through which our power of self-determination is exercised. What makes self-determination possible is that as a mode of exercising rationality action is quite distinctive. Intentional action, as we ordinarily understand it, is not any expression of prior motivation, but a special mode of exercising rationality in its own right—one which is distinct from and independent of any other.

Scanlon would certainly admit that action is a mode of exercising rationality—that in acting intentionally we are exercising a capacity for responding to justifications. But as we have seen he is committed to denying that what distinguishes action and separates it from the passive is the way that reason is involved in how we act. To perform an action is certainly not to exercise reason

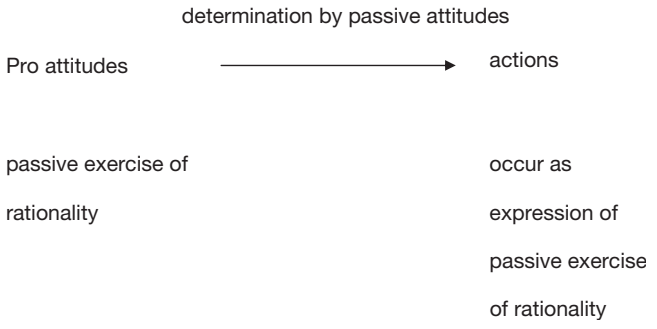


Fig. 5.2. Ethical rationalism

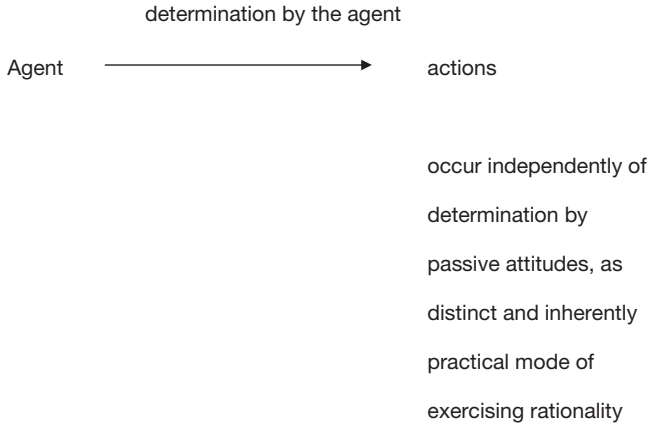


Fig. 5.3. Action as ethically distinctive

in any very special way, since, on his view, the rationality with which we act is no more than an expression of the rationality with which we are first passively motivated to act. The rationality of action is entirely parasitic on and derivative from the rationality of prior motivation that is passive. So in fact our capacity for rationality is exercised in the same way in our passive attitudes and in our actions. Action stands out only in its relation to these passive attitudes—as being their expression, as being what they have motivated us to do.

But on the *practical reason-based* theory of action I shall be defending, there is a clear distinction between action and any merely passive exercise of rationality. What makes action genuine action is not its being an expression of some prior exercise of rationality in passive motivation, but its being a distinctively practical, action-constitutive exercise of reason in its own right—an exercise of rationality

that begins in the will. Action can then occur quite independently of prior pro attitudes that are passive, determined to occur not by them but by ourselves.

8. ACTION AS A PRACTICAL MODE OF EXERCISING RATIONALITY

On any view, an action is a kind of event. To perform an action, to do something, is always to make a change. What distinguishes actions from other events or changes is one feature in particular—goal-directedness.

Goal-directedness involves more than an event's serving a function. For example the beating of a heart serves to circulate the blood and keep a body oxygenated. But the beating of a heart does not exemplify the goal-directedness we find in genuine agency. The goal-directedness that distinguishes agency involves, not simply an event's serving an end, but the event's being motivated by that end. An action is an event which has an object—an object at which the agent is directing the action's performance. The attainment of this object is what motivates the action's performance, being the goal at which the agent is aiming, and which he is using the action to attain. The goal may lie in some further outcome beyond the action. Or it may lie in the action's very performance; the action may be being performed for its own sake. This goal-directedness, this property of being done by its agent as a means to some end, is something which all actions have in common. And it distinguishes events which are actions from events which are not.

The voluntariness-based model understands goal-directedness as something determined and provided by prior attitudes that are passive, passed on from their contents to the voluntary actions that are their expression and effect. The practical reason-based model has an understanding of goal-directedness that is very different. It takes goal-directedness to arise as a distinctive mode of exercising rationality—not as an expression of prior motivations and rationality in passive form, but as a feature immediately arising in and internal to motivation that is active.

The practical reason-based model of action certainly says that to act intentionally is to pursue goals, and to pursue them on the basis of one's capacity for rationality. But it goes further, and maintains that to act intentionally just is to make a distinctive exercise of one's capacity for reason or rationality—one that is practical or action-constitutive because goal-directed. And this is something which one can either do competently and with due care, and so rationally; or incompetently and recklessly, and so irrationally. The model has room for agency which is irrational.

As we have seen, one can very well exercise one's capacity for rationality in ways that are not practical or action-constitutive at all. For example, there can be an exercise of one's capacity for rationality which is purely theoretical. Here

one responds to an object of one's thought simply by coming to believe that it is true. And the standards which determine the rationality of such a response are going to be correspondingly theoretical in nature. For one's response to be justified, and the belief to be formed rationally, the object, what is believed, must be sufficiently likely to be true.

But one can also exercise one's capacity for rationality in a way that is fully practical—that really is action-constitutive. One does this when one responds to an object of thought as one's goal—as something which the exercise of one's rationality, the very response to the object which one is making, is directed at attaining. For example, one can exercise one's reason practically in this way when, on the basis of an intention to do *A*, one voluntarily does something as a means to ensuring that *A* is done. In performing this voluntary action, one is exercising one's capacity for rationality. One's voluntary action is a response directed at an object, the doing of *A*—an object provided by the content of the intention which is motivating the action. And one is responding to this object in a rationally appraisable way. The response is one for or against making which there can be justifications, and which can be made more or less rationally. And the response is action-constitutive because the object responded to stands to the response as its goal—an object which the response is directed to attaining. What shows that the object is a goal? What shows this is the way the voluntary action's rationality depends on this object at which it is being directed. The mode of dependence is precisely that appropriate to something done in order to attain the object. For the action to be performed rationally, that object must be sufficiently desirable, and the action must also be sufficiently likely to attain the object.

This is a fully practical or action-constitutive mode of exercising reason. The agent is using the exercise of his capacity for rationality in order to attain a goal. He is responding to an object of thought—and doing so precisely in order to attain that object of thought and make it real. And so the rationality of the response depends both on the sufficient desirability of the object to be attained, and on the sufficient likelihood that the response really will attain it. But this fully practical mode of exercising reason is not to be found in voluntary action alone. It is found in non-voluntary intention-formation too—and in equal and unqualified measure. Such a practical mode of exercising reason is also found when one decides or forms an intention to do *A* rather than *B*. For such an event is again an object-directed exercise of one's capacity for rationality. It is a rationally appraisable response to an object—the object of the intention being formed. And again the object in question stands as a goal which one's response to it is being used to attain. For the rationality of the intention-formation is determined in just the same way by its object as in the case of the voluntary action—by reference both to the desirability of the object and also to the sufficient likelihood that the response to it in question, the intention-formation, will help attain it.

Compare deciding to do *A* with another event in which reason is also exercised non-theoretically—the event of forming a desire to do *A*. If, as the practical

reason-based model supposes, in decision-making reason is being exercised fully practically, in the sense of constituting fully intentional action, the same cannot be true of forming a desire. For while it is natural to assume that deciding or forming an intention to do *A* is an action—something which we deliberately do—it is equally natural to assume, as we have already seen, that forming a desire to do *A* is a passive occurrence. For we do not and cannot plausibly control directly what desires we form as we directly control how we decide to act. As we have seen, desires are states which are passive—which happen to us. Which is why any control which we may have over our desires is always indirect, being exercised manipulatively through the effects of prior actions on what we want—as when we employ exercise as a means of increasing our desire to eat. But if taking a particular decision, as something which we can control directly, without having to use such prior manipulation, is an intentional, goal-directed action, whereas forming a particular desire is not, wherein lies the difference? What makes intention-formation a case of action when desire-formation is merely a case of passion?

What is essential to the goal-directedness of a decision is the relation of the decision, as the formation of a content-bearing attitude, to the object which that attitude's content specifies. The decision is related to that object, namely the action or outcome decided upon, as to a goal to be attained through the decision's occurrence. And what establishes that relation of the decision to its object as a relation of means to end to be attained thereby, is the way in which reason governs decision-making. For a decision to be taken rationally, it is not enough for the object of the decision to be desirable. There must also be some sufficient chance that taking the decision will lead to the attainment of its object. Reason treats decisions as goal-directed exercises of rationality—as events of exercising reason which are being employed as means to attaining their objects, and so as actions. And if reason treats decisions as just such goal-directed actions, that must be what decisions are.

But what is true of decisions is not true of desire-formations. In forming a desire I am not yet doing anything in order to attain the desire's object. I am simply attracted to the object. I am responding to it simply as something really or apparently desirable. Which is why there is nothing irrational about coming to want something to happen while being very sure that if it happens, it will happen other than because one wants it to happen. We have and report such wants and desires all the time, and no one criticizes their rationality. Who do I want to win the Cup? It's England that I want to win the Cup—simply because, from my point of view, England's winning would be highly desirable. Perhaps England is my national team; or perhaps it's just that England's winning would put my boss in a good mood. Such considerations are quite enough to leave an England win a perfectly sensible thing for me to desire. I need not also suppose, what is obviously false, that my wanting England to win will actually help England to win. What I happen to want will have no effect at all on England's chances of

winning. But that does not matter. Since my desire for England to win is not something I am actually forming in order to get England to win, the fact that the actual outcome will be quite unaffected by the desire is quite irrelevant to the rationality of my forming it.

Sometimes it's not just that I want something to happen which will happen, if it does, independently of my wanting it to. Its happening independently can be part of what I want, and very much matter to the desirability of what I desire. Parents, for example, might reasonably want their grown-up children to do the right thing—but to do the right thing for themselves, and quite independently of the fact that their parents want them to do it. Suppose a parent is indeed sure that his grown-up children's actions are by now entirely beyond parental influence. What the children end up doing is sure to be causally independent of that parent's wants and attitudes. Given this belief, it is still both rational and natural for the parent very much to want his children to do the right thing—all the more so as this, if the child does it, will be its independent achievement. What, under the circumstances, the parent cannot rationally do is *decide* that his children will do the right thing. And that is precisely because reason governs the decision as a goal-directed action—as an exercise of rationality with a goal, namely ensuring the doing of the decision's object, what has been decided. For it is always and obviously irrational to employ an action as means towards an end when one is sure the means employed will have insufficient or no effect on the attainment of that end.

So the difference between the decision and the desire-formation is this. Taking a decision is an object-directed exercise of one's reason which is governed by reason as a goal-directed action—as a means employed towards an end, the end being that what is decided upon should occur. Hence a rationally taken decision must be a decision which is sufficiently likely to attain its object. Whereas forming a desire is an object-directed exercise of one's reason which is not governed by reason in the same way. In determining the rationality of a desire, we do not treat the object of that desire as a goal which the desire-formation is being used to attain. A rationally held desire need offer no prospect whatsoever of attaining its object. So while in decision-making, one is exercising reason practically or in a goal-directed and so agency-constitutive manner, in desire-formation the exercise of reason is not similarly practical.

Often, when we take decisions between options, we assume that each possible decision is as likely to attain its goal, performance of the action decided upon, as any other. But this need not be so. Consider a case which I have discussed in greater detail elsewhere.[13] I might presently prefer and want most of all to do *A* rather than *B* in the future—and given the risks involved in doing *B*, such a want may be very sensible. Perhaps I am a professional stuntman—and *B* is at

[13] For a more detailed version of this example, see my 'Purposive Intending', *Mind* (1991), and my *The Psychology of Freedom* (Cambridge: Cambridge University Press, 1996), ch. 8.

some future date attempting a particularly dangerous stunt, while *A* is prudently refraining. But if, because of a likely intervening increase in my recklessness, there is at least a significant chance (if no more) of my subsequently abandoning any decision I now took to do *A* and of doing risky *B* anyway; and if, from the point of view of planning other matters, there is enough of a premium on settling now as surely as possible which of *A* and *B* I shall be doing—then notwithstanding the rationality of my present preference for doing *A*, it might be more sensible for me now to decide to do *B*. For that decision would be sure to attain its goal, which is causing me to act as decided. The point and purpose of a decision is to cause me to execute it. Which is why a decision that would be too ineffective at attaining this end can be a decision that I have reason not to take.

How likely a decision is to attain its object matters, then, to the rationality of a decision—as the likelihood of any action's attaining its goal is relevant to the rationality of that action. But how likely a desire or preference is to attain its object is not similarly relevant to the rationality of forming that attitude. And so, precisely because of this, it can be rational to want and prefer things to happen which one cannot rationally decide should happen.

This story told of why decisions are actions has nothing to do with voluntariness or with the decision's being an expression of prior pro attitudes to its being taken. It is the goal-directed exercise of rationality that is constitutive of the decision event's being an intentional action. But that goal-direction is not, as on the voluntariness-based model, a function of the content of a prior passive motivational cause. It is instead a function of the content of the intention being formed and of the mode of direction which the decision constitutes towards the object which that content presents.

Which is why, on the practical reason-based model, any role played in the motivation of action by passive pro attitudes such as desires, is entirely contingent. For such desires or other passions play no role in constituting the character of what the agent does as goal-directed action. Our actions may indeed frequently or even typically be influenced by prior desires. But the existence and degree of that influence has to be determined on a case-by-case basis. Motivation by the passive is no longer essential to the very character of action.[14]

Action, then, is no longer the mere expression of a mode of exercising rationality that is passive—that immediately takes place in the formation of

[14] There is a fundamental division between the action theory assumed by much modern English-language ethics and the action theory that underpinned the moral philosophy both of the scholastics and of Kant. If we use 'desire' as a catch-all to pick out any motivation that precedes agency as passive, then there is a modern tendency to assume that all action is an effect and expression of desire. For such is a direct consequence of the voluntariness-based model of action. But for the scholastics and for Kant there is no difficulty whatsoever in the possibility of action that is unmotivated by desire, and that occurs quite independently of any passion or passive motivation. This has nothing to do with any extravagance in scholastic or Kantian metaphysics, but is, I would argue, a simple consequence of a very different way of understanding goal-directedness—i.e. practical reason-based, and not voluntariness-based.

motivations that are not the agent's intentional doing—but a distinct and distinctive mode of exercising rationality in its own right. And that allows action to have the independence of the passive required for real self-determination to be possible. Not only that but our freedom can be exercised as we ordinarily suppose it to be—in and through our decision-making, and so in our fundamental response to morality and its demands. That allows us to be genuinely responsible for whether we regard or disregard morality—a moral responsibility that can at the same time be what we ordinarily take it to be, a responsibility for how we act. If self-determination is a power that goes beyond any mere exercise of reason, its reality and moral significance still depends on the peculiar and special way that rationality is involved in human action.[15]

[15] This account of self-determination and action and their relation is developed further in my forthcoming *The Ethics of Action: Self-Determination*. That volume will centre on the philosophy of mind and action. A companion volume centering on moral and legal theory, *The Ethics of Action: Normativity* will develop the account of blame, responsibility, and obligation. The practical reason-based model of action was the model assumed by medieval scholasticism and by much early modern natural law theory, both Catholic and Protestant. For historical accounts of the model, and of Hobbes's opposition to it and of the profound impact of this opposition on subsequent moral theory, see my 'Moral Obligation', in A. O'Hear (ed.), *Modern Moral Philosophy* (Cambridge: Cambridge University Press, 2004); my 'Suarez, Hobbes and the Scholastic Tradition in Action Theory', in M. Stone and T. Pink (eds.), *The Will and Human Action: From Antiquity to the Present Day* (London: Routledge, 2004); my 'Action, Will and Law in Late Scholasticism', in J. Kraye and R. Saarinen (eds.), *Moral Philosophy on the Threshold of Modernity* (New York: Springer, 2005); and my 'Natural Law and the Theory of Obligation', in S. Heinemaa and M. Reuter (eds.), *Philosophy and Psychology* (New York: Springer, forthcoming).

6

Freedom and Practical Judgement

David Owens

Human beings can choose what to do. Human beings can also act freely. Many writers think the one fact helps to explain the other, that if spiders cannot act freely that is because they cannot choose what to do. True most human actions are performed without first becoming the topic of choice, as when you turn the pages of this chapter. Nevertheless, once you have started reading, you do so freely because you have the capacity to choose whether to carry on.

I agree that our capacity for free action depends upon our capacity for choice. In particular, it depends upon our ability to arrive at a practical judgement, a judgement about what to do and to implement that judgement in action. But to make a case for this, we must assure ourselves that practical judgement is under our control in just the sense that action and intention are under our control, for our ability to control our practical judgements can't be the source of our ability to act freely unless we control our practical judgements as we control our actions and intentions. That should make us suspicious of the idea that practical judgement is a kind of belief for, it is generally agreed, we don't control the formation of our beliefs as we control our actions and intentions.

In this chapter, I'll first ask what a practical judgement could be if not a belief. Then I'll argue that we have a capacity to make and enforce practical judgements, so understood, whenever we are acting freely. Finally, I'll seek to establish that the making of a practical judgement is free in the very sense in which actions and intention formation are free and so can indeed be the source of our practical freedom.

1. WHAT IS PRACTICAL JUDGEMENT?

Practical judgement is a judgement about what to do. Beginning here many authors move on to say that a practical judgement is a judgement about what I (or we) ought to (or should) do and they come to rest with the claim that practical judgement is a judgement about what we have most reason to do, or

what it would be best to do all things considered. Although these transitions are frequently made, and though I shall make them here, they shouldn't pass without comment. Some creatures might be capable of thinking about what to do without deploying normative concepts like 'should' and 'ought'. And we who can deploy such concepts might decide what we should or ought to do without thinking about reasons, values or what was best. Nevertheless, to facilitate discussion I shall here assume that when a rational human agent thinks about what to do, they are thinking about what action reason recommends.[1]

The idea that practical judgement is a form of belief has been disputed by some who worry that a mere belief would not have the required motivational impact (e.g. Nagel 1970: 65). But many others are happy with the idea that a belief can play the motivational role of practical judgement. Scanlon tells us that we can 'explain the intrapersonal rational significance of judgements about reasons for action . . . without supposing that those judgements are anything other than beliefs' (Scanlon 2003: 19). Broome also sees no difficulty here 'We often deliberate in order to arrive at a normative belief about what to do, and the point of our deliberation is ultimately to bring us to a decision—the forming of an intention' (Broome 2001: 180).

In this chapter, I shall assume that Scanlon and Broome have right on their side in maintaining that there are truths about reasons which we can know, the knowledge of which can move us to action. It might seem a short step from this to the view that practical judgement is 'just another belief' (Arpaly 2003: 61). For example, having told us that he is 'strongly drawn to a cognitivist understanding of . . . practical judgements' Scanlon adds 'They strike me as the kind of things that can be true and their acceptance seems to be a matter of belief' (Scanlon 2003: 7). But mightn't practical judgements be 'the kind of things that can be true' without also being the kind of things whose acceptance is a matter of *belief*? I want to argue that, on at least one familiar construal of the term 'belief', practical judgements are not beliefs.

What are the connotations of this familiar notion of belief? I shall distinguish three. First, belief is governed by a norm of truth. Second, a belief motivates action on that belief. Third, belief is governed by a norm of knowledge. It is this third connotation that distinguishes belief from what I call practical judgement. I shall review these features in order.

Belief is subject to a norm of truth in that it is correct to believe that p only if p is true. This distinguishes belief from states like desire and intention that can be neither true nor false. It also distinguishes belief from activities like imagining, supposing, and hypothesizing which can be more or less truthful (Shah and Velleman 2005: 499). For example, when I imagine that the figure is bisected, my imagining is false but not thereby incorrect. I am not wrong to

[1] For more on the presuppositions of deliberation both theoretical and practical, see Owens (forthcoming).

make that supposition; there is no standard normative for imaginings which it violates simply in virtue of being false. By contrast a false belief is *ipso facto* an incorrect belief. To believe something false is to believe wrongly; it is to make a mistake.

Perhaps some writers move from the premise that practical judgements are subject to a truth norm to the conclusion that they are beliefs because they assume that only belief is subject to a truth norm. Yet guessing that *p* and suspecting that *p* are both incorrect if false. If your guess or suspicion is false, you have guessed or suspected wrongly, you have made an error.[2] In this respect, both guessing and suspecting differ from imagining, supposing, and hypothesizing. A scientist employed to pursue a certain line of research might put forward an hypothesis or adopt a supposition for the sake of argument and he might be quite correct to do so even though his hypothesis or supposition is untrue.

Those who note that belief is subject to a truth norm often go on to say that beliefs must 'aim at the truth' and that this is why beliefs cannot be adopted at will regardless of evidence. Do guessing and suspecting aim at the truth in the same way? My own view is that this talk of 'aiming' adds little to our understanding of belief and so I don't propose to investigate whether it can help us with guessing and suspicion.[3] I observe only that, whilst we can suppose or hypothesize that Bill Clinton is a woman just because we have been offered a large financial incentive for so doing, we can neither believe nor suspect nor guess this of Clinton simply to get such a reward (though we can *say* 'I guess'). On that point, believing, suspecting, and guessing are entirely at one. I leave it open how this similarity is to be explained and, in particular, whether it can be explained by reference to the truth norm.

So how are we to distinguish belief from other states governed by a norm of truth? A second connotation of 'belief' is practical: someone who believes that *p* has a default rational entitlement and a prima facie rational obligation to act as if *p* is true (that is to do what would be reasonable on that supposition).[4] Many will think this claim too weak but I shall postpone that particular discussion for a few pages. Even this weak claim is enough to differentiate belief from both guessing and suspecting. It can be reasonable to act as if your guesses or your suspicions

[2] In other respects, guesses and suspicions are quite different. Guessing is a mental action, an action sometimes but not always expressed in a speech act (Owens 2003: 289–96). Suspicion is more like belief—you come to suspect, you don't decide to suspect—but suspicion does not require the same evidential warrant as belief (Greenspan 1988: 90–1). (Note there is a factive use of 'guess' on which you fail to guess someone's age if you get it wrong but there is another usage, which I here adopt, on which you can guess incorrectly.)

[3] I think it will help more with guessing than with suspicion. For more on the aim of a guess, see Owens (2003).

[4] Broome (2001: 181–2) argues that one need not have any reason to act on one's beliefs (e.g. if they are false). I shan't discuss Broome's view here. What I presuppose is something Broome will allow, namely that there is (in a way to be qualified) some irrationality in failing to act as if your beliefs are true.

are true but there is no default entitlement or prima facie obligation to do so. Whether this is reasonable will depend on further facts about the situation.

Have we said enough to differentiate belief from other mental phenomena? I doubt it. There is a further connotation of 'belief' which makes the cut between belief and practical judgement. Belief is not just subject to a norm of truth, it is also subject to a norm of knowledge: if what you believe is something you don't know then your believing it is incorrect. To show someone that they don't know that *p* (perhaps because they are not justified in believing it) is to show them that they are also mistaken in believing it. To think yourself right to believe that *p*, you can't think you are ignorant of its truth, anymore than you can think *p* false (Owens 2000: 37). It might be an accident that you are right to guess that *p* but it won't be an accident that you are right to believe that *p*.[5]

I don't deny that there are uses of the term 'belief' which would not sustain this connection with knowledge; I assert only that there is a notion of belief which does, a notion familiar from both epistemology and ordinary life. It is not clear to me whether the writers I quoted at the outset are employing this epistemic notion of belief in formulating their claims about practical judgement. But, as we shall see, clarity on this point is crucial to our understanding of practical freedom.

Some maintain that this third feature of belief is not really independent of the other two. This is wrong. One's guess or one's suspicion can be perfectly correct because true even though one fails to know that it is true. Guessing and suspecting demonstrate that something can be subject to the truth norm without being subject to the knowledge norm. Furthermore, we can at least conceive of a mental phenomenon which is governed by the truth norm but is unlike guessing, etc., in that one *does* have a default entitlement (obligation) to behave as if it were true even when one is not in any position to claim knowledge of its truth. This possible phenomenon I shall call a practical judgement.[6] In this chapter, 'practical judgement' refers to something which is like belief in that it should dispose us to behave as if it is true yet unlike belief in that it is subject to the truth but not to the knowledge norm.

It is one thing to point out that a phenomenon like practical judgement might exist, it is quite another to establish a real theoretical need for it. In the next section, we shall see that practical deliberation often terminates in something both subject to a truth norm and action motivating but which does not constitute a belief.

[5] In Gettier cases you have committed an error even though your belief is both reasonable and true.

[6] 'Judgement' is sometimes used to refer either to a kind of internal assertion which manifests belief or else to a cognitive action which is a precursor of belief. For example, both Shah and Velleman (2005: 503) and Peacocke (1998: 88) use the term 'judgement' to denote a mental act 'aimed at the truth' which concludes theoretical deliberation and so (in a rational person) is a precursor of a belief. I doubt the existence of such 'theoretical' judgements. It may be that practical judgement is a mental action but this point needs careful handling and I place no weight on it here.

2. JUDGEMENT AND IGNORANCE

I am assuming that the practical deliberator is seeking to determine what to do by determining what he ought to do, by settling what action reason recommends. But we shouldn't assume that practical deliberation concludes only when the deliberator thinks they *know* the answer. A question like 'What ought I to do?' may be answered with a guess or by the formation of a suspicion, at least where knowledge of the answer is unavailable. And, I want to suggest, a deliberator can make a practical judgement, even where he knows he has yet to dispel his ignorance about what to do.

I am trying to settle on a suitable restaurant for our anniversary dinner. We live in a large city and there are many to choose from. In the case of some, I take myself to know that there is nothing to choose between them. If those restaurants are also the best, I might simply pick one of them at random. I thereby decide *to* dine at a certain restaurant (i.e. form the intention to dine there) without deciding *that* we ought to dine there: if I went to the one I didn't choose by mistake, I wouldn't be doing anything I judged I ought not to do. On this point, my action is not guided by my judgement as to which option is best.

There are other cases in which my action *is* guided by my judgement, even though I am ignorant of which option is best. Suppose that I neither take myself to know which of the more attractive restaurants is best, nor that they are equally good. In fact, I'm pretty sure there are significant differences between them which further investigation would reveal and which might well affect my choice. But one can only spend so much time choosing a restaurant, even for an anniversary dinner, so I make reservations at the restaurant which seems the best on present showing. Here I judge that we ought to go to this restaurant and I do so on the basis of my beliefs about desirable features of the restaurant but I wouldn't claim to know that this restaurant was the best or the most desirable or the one reason favours. That I know I'm currently in no position to know.[7]

My way of proceeding is unobjectionable. What I am doing is registering in my practical deliberations limitations on the process of deliberation itself. These limitations are hardly shameful. On the contrary, they are constraints under which all finite creatures labour and it makes perfect sense to take them into account when determining when to make up your mind. Because of them rational practical deliberation often concludes with (and does not merely stop at) a proposition that the deliberator wouldn't claim to know.

[7] Holton (2006: 7–9) observes that we often make choices even though (*a*) we believe the options to be both significantly different and commensurable and (*b*) we know that we are in no position to know which option is best. Holton describes these as cases of 'choice in the absence of conscious judgement' (p. 9). He says this is because he equates conscious judgement with conscious belief (p. 6).

Can we accommodate this point by instead relaxing our assumption that practical deliberation aims to discover what we have most reason to do? Perhaps I need not take myself to know that this restaurant is the best one before concluding that I ought to dine there but don't I at least take myself to know that it is good enough for my purposes, that there is no decisive objection to it, that given the limited time and effort I can expend on the matter, this is the option I ought to go for? If so the judgement which concludes deliberation may be a normative belief after all.

Sometimes we can indeed know that a restaurant is good enough in this sense, without being in any position to know that it is the best. But, on other occasions, the issue as to whether a restaurant is 'good enough' might be hardly less difficult to settle. For example, whether it is 'good enough' might depend on whether there is a nearby restaurant which is known to be better by my partner. Perhaps I have time to eliminate this possibility (by obliquely questioning my partner, etc.). In that case I may know that the restaurant is good enough without necessarily knowing that it is the best all round. But suppose I don't have the chance to rule this possibility out. I might still judge that we ought to go ahead and eat there.

Nor will it help to build the constraints on the deliberative process explicitly into the content of my practical judgement. I *might* be in a position to know that, given my limited deliberative resources, the thing I ought to do is to settle on the option which currently seems best (the one which I suspect to be best). On the other hand, I might not. I might only be in a position to make the educated guess that I ought to make a judgement at this stage rather than holding out for more information. When, on the basis of this guess, I judge that I ought to conclude deliberation now, this is no mere stab in the dark. But equally I wouldn't claim to know that I ought to do this, for I know it is a real possibility that a bit more deliberative effort would yield results and so forth. Nevertheless I judge that what I ought to do is to settle on this restaurant.

Someone may wonder whether these decisions about what one ought to do are genuine judgements. Are they not decisions *to* rather than decisions *that*? It is certainly true that the point of making such a judgement is not merely to evaluate the options but actually to get yourself to act. In that the judgement is like an intention. And at least where future action is in question, the skill of making practical judgements just is the skill of drafting sensible intentions. Still this should not blind us to some obvious differences between them.

First, intentions may be reasonable or unreasonable but it is doubtful whether they are correct or incorrect and they are certainly not assessed for truth. By contrast the judgement that I ought to φ is evaluated not just as reasonable or unreasonable but also as correct or incorrect and its correctness depends on its truth value. Second, even though a good practical judge will opt only for feasible intentions, the fact that he judges that he ought to φ no more guarantees that he will form the intention to φ than it guarantees that he will φ when the time comes. Both lapses are possible.

To sum up, in making a practical judgement I need not form a belief about which option is best or even good enough. I need not form a belief about what I ought to do. I can be living in ignorance on all these points.[8] But I may still conclude my deliberation with a *judgement* about what I ought to do and sensibly implement that judgement in decision and action. This practical judgement shares two of the features I took to be characteristic of belief: it is correct only if true and it should move me to act as if it is true. But it is not a belief for though I think myself entitled to make it, I lay no claim to knowledge of its truth.

From what has been said so far, practical judgement looks rather like the better sort of guess but the comparison is misleading. A practical judgement is not, like an educated guess, just a way of dealing with ignorance. If it were, it could hardly be the source of our practical freedom for we don't think ourselves free only where we are ignorant of what to do. In the next section, I shall argue that practical judgement can actually countermand belief, both normative and non-normative, for it sometimes makes sense to judge that you ought to act as if what you *know* to be so isn't so. Such a practical judgement won't be based on a guess since you can't (reasonably) guess to be false what you know to be true.

3. JUDGEMENT AND KNOWLEDGE

How could it ever be sensible for us not to act on what we know? To answer this question, we must ask why knowledge matters to us. I have argued that belief is not the only action motivating state subject to a truth norm. If so, why do we value knowledge? Why does it matter to us whether we can *know* the answer rather than make an educated guess at it?[9] We want to be right and someone who knows is more likely to be right than someone who doesn't but that fact alone won't explain the importance of the boundary between knowledge and ignorance. We are much more likely to be right about some of the things we take ourselves to know than about others; furthermore, we are almost as likely to be right about some of the things we don't take ourselves to know as about many of the things we do.

Since belief embodies a claim to knowledge, to ask about the value of knowledge is to ask why we want beliefs, why we want to satisfy our curiosity, to

[8] In previous work, I left it open whether a rational person must know their reasons, whether they can be in ignorance of what rationality requires of them (Owens 2000: 15–16). I now think that whilst a rational person must act on their view of what reason recommends, one can take a view about what reason recommends without taking oneself to know what reason recommends (Williamson 2000: 180). In particular, ignorance about what one should do cannot always be dispelled by knowledge of some principle of decision-making under uncertainty.

[9] Those interested in the value of knowledge often ask why knowledge is more valued than mere true belief. This is not the best way of putting the question if belief is subject to a norm of knowledge.

remove our doubts, to resolve our uncertainties. Guesses and suspicions, however well founded, will never suffice, though they may carry just the information we need. One popular answer presupposes that if someone is entitled to believe that *p* then it must be sensible for them to act as if *p* is true. Things we know are, it is said, things we can (and must) assume to be true for the purposes of action.[10] On this view, 'the importance of the concept of knowledge' resides in the fact that 'it sets a meaningful lower bound on strength of epistemic position: your epistemic position regarding p must be strong enough to make it rational for you to act as if p is true' (Fantl and McGrath 2007: 581).

In discussing the motivational role of belief, I agreed that things we believe do play a special role in our practical deliberations. But I doubt this role is best captured by making unqualified claims like the following: 'S is justified in believing that p only if S is rational to act as if p' (Fantl and McGrath 2002: 78). First, such claims postulate too tight a connection between belief and rational agency. It is not always rational to act as if *p* because you know that *p*. In particular, this is not rational in some cases where (*a*) the costs of acting as if *p* should *p* turn out to be false are substantial or (*b*) the benefits of acting as if not *p* should not *p* turn out to be true are substantial.

Second, such claims focus on the connection between belief and agency to the exclusion of other equally significant features of belief. Our convictions play a crucial role in our emotional psychology. Someone can be angry at the fact that *p*, or proud of it, or grateful for it, only if they know that *p* (Gordon 1987: 47–9). Often we want to know whether *p* in order to fix our emotional bearings, to avoid having our feelings baffled by ignorance; in eliminating uncertainty we learn how to feel as well as how to act. Sometimes we would prefer to stick with those emotions—hopes and fears—that presuppose uncertainty rather than learn the truth. Still the boundary between knowledge and ignorance retains its emotional significance.

In the light of these points, I propose the following as a partial account of knowledge—to know that *p* is at least to be justified in using *p* (i) as a default assumption in your practical reasoning and (ii) to inform whatever cognitive processes guide the higher reaches of our emotional lives.[11] A default assumption is one that you can rely on when you have no specific reason not to. Memory is full of beliefs we depend on in this way and cognition could not take its current form unless we were entitled to this reliance. Such default assumptions are indeed crucial to all practical reasoning. Even where we end up making an educated guess, we feel entitled to act on that guess (i.e. to judge that we ought to) only because of an assumed background knowledge of the situation. But it is

[10] Something like this is presupposed by many of those who argue that whether we know that *p* depends, in part, upon our practical interests in the truth of *p* (e.g. Hawthorne 2004: 173–81).

[11] I say 'at least' because propositions we take ourselves to know may also play a distinctive role in our theoretical reasoning, e.g. they may serve as our evidence (Williamson 2000: 203–7).

not *always* advisable to act on these assumptions. Nor is the advisability of acting on these assumptions the only thing which makes them worth having. On my view, knowledge is valued more than an educated guess because a proposition known forms part of that framework of default assumptions which we need to conduct both our cognitive and our emotional lives.

Let's test a purely agency-based view of the value of knowledge against my own proposal by considering some examples. I have parked my car on the street outside, taking the amount of care a reasonably conscientious citizen would to park legally. When I enter the house, my partner informs me that the police have been ticketing the street this week. Before being told of this, I took myself to know that my car was parked legally, that is, I took myself to have evidence sufficient to justify my believing this. Hearing my partner's words, I reluctantly go out and recheck the position of my car and the relevant parking notices. Is this an implicit admission that I no longer know that my car is legally parked, at least until I have completed the checks, because my belief is no longer justified?[12]

Our linguistic intuitions here seem inconclusive. Being reluctant to check, I might say to my partner 'I know I'm properly parked' and they might reply 'Yes I agree but it is still worth checking'. That sounds as if my partner is agreeing that I do know and thus agreeing that they can learn from me how the car is parked whilst also suggesting that here it might not be sensible to act on our knowledge since the costs of being wrong on this point are substantial and the check can easily be made. Is my partner merely being polite? Or are they observing quite sincerely that this is one of those cases where practical judgement should countermand the motivational effects of a default background assumption to which I am still perfectly entitled?

Suppose my partner instead says 'But do you *really* know the car is properly parked?' Now it sounds as if I *am* being invited to abandon my belief and to do so because it has become unjustified. But this isn't the only interpretation available. Perhaps my partner is highlighting the possibility that my belief is false, a possibility that would deprive me of knowledge even if my belief were still fully justified, a possibility on which I must now focus for practical rather than epistemic reasons. My partner might be seeking to influence my practical judgement without thereby changing my convictions.

Our car example is one in which I am sensible not to act on my (well-founded) assumption that *p* because the costs of being wrong are substantial. Similar issues are raised by cases in which I risk missing out on a considerable, though unlikely, benefit if I act on my default assumption. Suppose someone offers to pay me ten million pounds in return for a stake of ten pence if it turns out that I was not brought up a Catholic. I know that I was brought up a Catholic and much of the rest of what I know about myself would make little sense were I not. Nevertheless

[12] A similar example is used in support of this conclusion by Fantl and McGrath (2007: 560).

I might reasonably accept the bet (Hawthorne 2004: 176).[13] Can the mere fact that I have been offered this bet render one of my most well-founded convictions unjustified? It would sound odd for me to confess ignorance of which religion I was raised in because it is silly to miss out on this bet. On the other hand, as I place the bet, I might say to myself 'Well I guess I *might* be wrong about which religion I was born into' and then it would be slightly strange to add 'but I *do* still know'.

There is something awkward about describing yourself as acting on an assumption which you know to be false but the awarkwardness is, I reckon, just the awkwardness of explicitly acknowledging the possibility of error in a context in which you also claim knowledge. 'I know that p though I'm not absolutely sure that p' jars, as does 'I know that p though I might be wrong' but we shouldn't infer that one who takes themselves to know cannot sensibly acknowledge their own fallibility. On the contrary, rationality requires such an acknowledgement from us all and rationality permits us, on occasion, to act on it by not assuming in our practical deliberations things we take ourselves to know.

On my view belief and practical judgement are each, in their different ways, fundamental to our lives as agents. In one way belief is more fundamental. No agent could get by on judgement and conjecture alone. Without that background of default assumptions, one could make neither guesses nor practical judgements. On the other hand, it is true of (virtually?) any belief that a rational agent has the capacity to countermand its motivational effects by judging that it would be right not to act on it.

As already noted, belief and knowledge have a life outside our practical deliberations, underwriting a rich emotional psychology. For example, I can feel proud that I was raised a Catholic, or ashamed for that matter, only if I know that I was raised a Catholic. If I don't think I know this, whilst I can think myself entitled to entertain hopes or fears on the matter, I can't think pride or shame would be in place. Yet I won't come to think pride or shame impossible just because I have been offered the bet. Pride and shame need not come and go in response to such offers, rather they are part of a more permanent background, dependent on relatively stable convictions which structure our emotional lives as well as supplying default assumptions for practical deliberation.

Now imagine that I am rather proud to own such a fine car, the very car I parked outside the house. On this occasion my partner informs me some time after I have arrived in the house that the police are confiscating illegally parked cars and were doing so in our street only last week. Must I cease to take pride in my car until I have checked that no such confiscation has taken place? Should I be gripped by fear for my social status? A sober person would rather judge that it is sensible to check how the car is parked and then calmly leave the house, convinced that their car is still there. Such conviction would be misplaced had

[13] I might reasonably accept *any* such bet but not *all* such bets.

they parked the car less carefully but it isn't misplaced simply because they judge that they should check.

What is true of pride (or shame if I feel bad about driving a status symbol) applies equally to anger, embarrassment, sorrow, joy, gratitude, disappointment, disgust, and much of the rest of our emotional lives. Behind these attitudes and reactions lie certain default assumptions about how the world is, assumptions which shape what we do, think, and feel in a wide range of contexts. If this cognitive background is to be so widely available, it can't be tied too closely to any one context. Even as I walk out the door to check my car, I may be using the default assumption that I have a car in various inferences (e.g. in thought about the best way to avoid tomorrow's traffic jams) as well as to take my emotional bearings. Why deny that I can be entitled to believe it, just because I am also confirming its truth?[14]

Several authors, myself included, have argued that the level of evidence required to justify a given belief depends, in part, on the needs and interests of the believer (Owens 2000: 24–7). And this raises the prospect that different subjects confronted by the same evidence for *p* may find themselves in a rather different epistemic situation with regard to *p*. Where this happens, I would maintain, it happens because of relatively permanent and pervasive differences between people, e.g. differences in social role, intellectual interests, or long-term personal relationships.[15] Our convictions are multi-purpose and changes or variations in what would justify them make sense only where a range of these purposes are affected by the relevant factors. Such variations are not brought about by transient changes in the stakes riding on particular issues.

For example, if someone is a close relative of mine, the amount of evidence I require before I begin to doubt their honesty is rather different from that required by a stranger. And if bird spotting is my main passion in life, I may not feel able to believe that a willow warbler has appeared for the first time in California even on the basis of several reported sightings, whilst it would be neurotic for the average newspaper reader to demand as much. In both cases, because it matters so much more and in so many ways for me to be right about this sort of thing, I should hold myself to higher standards in forming beliefs about it.[16]

[14] Similarly for testimony. I shouldn't be prevented from dipping into the fund of common knowledge by some ephemeral circumstance. I would be so prevented if I couldn't now learn that *p* from you just because it would not be sensible for me to act on this knowledge at this very moment.

[15] I am less confident of the existence of relevant variations between believers than I am of the proposition that the required level of evidence is fixed, in part, by the needs and interests of believers as such. That conclusion can be established simply by asking what else could fix this level: evidence certainly can't. Fantl and McGrath (2002: 71 and 87–8) complain that this simple argument against evidentialism tells us little about how the relevant level of evidence is fixed. I agree but this throws no doubt on its soundness as an argument against evidentialism.

[16] My account of the role of belief in practical deliberation is, in some respects, similar to that offered in Bratman (1999: 15–34). Bratman regards belief as providing a 'default cognitive background' for practical deliberation whilst maintaining that we may 'posit' things we don't believe

The points made in this section apply as much to beliefs about what I should do as to the non-normative beliefs on which they are based (*pace* Fantl and McGrath 2007: 571–4). Since I know my car is correctly parked, I know I would be wasting my time going to check it and so I know that I ought not to check it.[17] Yet I judge that I ought to check it. Of course, I make this judgement in the light of what I know about the (remote) chance of it being parked illegally and the trouble involved in checking. But I screen off the knowledge that it is in fact correctly parked in judging that I ought to check. Once I have checked and found the car to be legally parked, I will admit that my belief was right and my judgement was wrong: it wasn't true that I ought to have checked. Nevertheless both belief and judgement were perfectly reasonable.[18]

I conclude that though people have a rational entitlement and obligation to act as if their beliefs are true, this requirement is defeasible. Does the same apply to practical judgement itself? Is the rational requirement to act as if one's practical judgement is true also defeasible? I think so. Suppose I judge that I ought not to back this inventor and finance the production of his self-cleaning shirt. Then you offer me a bet asking for a small stake for a large reward should my practical judgement turn out to be false. Even if I am very confident of my practical judgement, the reasonable thing might well be to accept the bet. Here I don't abandon my practical judgement, indeed I act on it in that I allow it to govern my investment behaviour but I don't act as if it is true when accepting the bet. There is no failure of rational self-control here because I am behaving in accordance with my higher-order judgement that I ought not to act as if my first-order judgement is true. In a rational agent, practical judgement can countermand the motivational force of both belief and practical judgement.

4. PRACTICAL FREEDOM

Let's return to the question with which I began: how does our capacity for choice underwrite our practical freedom? The initial worry that action could not be free were our actions determined by our beliefs was met with the claim

or 'bracket' things we do when deliberating about particular issues (p. 29). For Bratman these positings and bracketings, unlike beliefs, are mental acts. Bratman's account differs from mine in that (*a*) he confines himself to belief's role in practical deliberation, (*b*) he asserts that reasons for belief are purely evidential and (*c*) he does not discuss knowledge (or freedom).

[17] Where change is likely the maintenance of knowledge requires periodic checking. This is not such a case: neither the position of your car nor the parking regulations are likely to change.

[18] Given that I have argued that belief and practical judgement are subject to a norm of truth, can one reasonably believe that *p* and judge that not-*p*? Can a rational person (knowingly) tolerate a situation in which they can't possibly be obeying both of the relevant norms? I think so. Several authors have noted that where one discovers an inconsistency amongst one's beliefs but can't tell which of the relevant beliefs is false, it may be reasonable to settle for inconsistent beliefs. This shows that the relationship between norms of correctness and norms of rationality is rather complex.

that we have an independent capacity to control our actions by means of our practical judgement. But this is no advance unless we are freer in making practical judgements than in forming beliefs. Is practical judgement any better suited to be the source of our freedom? First, I'll say what a mental phenomenon must be like in order to be the source of our freedom and then I'll argue that practical judgement alone satisfies that requirement.[19]

I'm exploring theories of free action according to which our freedom consists in our ability to use a certain psychological instrument to control our agency. Call this instrument *choice*. I am assuming that our freedom of action depends on a prior freedom of choice. But what is choice? Which bit of our psychology constitutes our choice? Is it a *belief* about our reasons, or else a *judgement* about our reasons, or else an act of *will* based on this belief or this judgement?

At the outset, I suggested that the psychological instrument of our practical freedom must satisfy a certain condition, namely that we be able to control that instrument as we control our actions by way of it—the source of our freedom of action must be free in just the way that action itself is free. In my view, practical judgement is the source of our freedom of action because practical judgement is under the control of practical judgement in just the way that action (and intention) are under the control of practical judgement. Where an action is free, this is so because (i) one can control the action (or intention) by making a practical judgement and (ii) one can control whether one makes the practical judgement by making a practical judgement.

Some maintain that action is not truly free unless every determinant of this action is itself freely chosen. It is an open question whether this demand can ever be satisfied or even coherently stated. But there is another thought which might be what lies behind at least some people's attraction to the impossible demand:

> *Constraint*: If one has freedom of action because one has freedom of choice, choice must control choice in the way that choice controls action.

What *Constraint* requires is that the regress of control terminate in a type of mental phenomenon which controls itself in just the way that it controls action. Note that a form of choice might satisfy *Constraint* even if such choices were entirely determined by factors (our upbringing, social environment, etc.) which we did not choose.

Among the candidates for the psychological instrument of our self-control, only practical judgement satisfies *Constraint*, or so at least I shall urge. I reject

[19] This chapter is about the psychology rather than the metaphysics of freedom. It asks what psychological states are distinctive of free agents. Other psychologies of freedom award the palm to higher-order desires or normative beliefs or yet other mental phenomena. I shall not commit myself on what sort of capacity a free agent must have to do, decide, or judge otherwise than he did. This question would be the focus of any inquiry into the metaphysics of freedom. In particular I take no stand on whether freedom as I understand it is consistent with determinism.

the idea that (normative) beliefs are the source of our practical freedom because one can't control those beliefs by forming beliefs about what one should believe. I also maintain that, though intention controls action, intention does not (in the same sense) control itself: the will is not subject to the will. So, I shall conclude, free choice consists in the making of a practical judgement.

Let's begin with the will. Voluntarists hold that what makes us free is our possession of a will, of a capacity to control our actions by forming intentions. For voluntarists, what makes our actions free is the fact that they are subject to our will (i.e. to our intentions). What does it mean to say that an action is subject to our will, that we can perform it 'at will'? Elsewhere, I have argued that something is subject to our will when we have the capacity to bring it about simply because its occurrence seems desirable to us. Many actions are subject to our will because we can bring them about for this reason. But, in that sense at least, the will is not itself subject to the will, for we can't form an intention to φ simply because it seems desirable to form that intention and regardless of whether φ-ing would itself be desirable (Owens 2000: 78–82). In fact our will is no more subject to our will than our beliefs are subject our will. We can no more form intentions 'at will' (i.e. form whatever intention would be most desirable regardless of the apparent value of what is intended) than we can believe 'at will' (i.e. form whatever belief would be most desirable regardless of the apparent truth of what is believed).

So the will fails to satisfy *Constraint* since the will does not control itself in the way that the will controls action. Nevertheless, the voluntarist is right to observe (*a*) that when we form intentions we are free as we are not when we form beliefs and (*b*) that nothing can be the source of our practical freedom unless it is itself free in the way in which the will is free. Even if our practical freedom does not have its source in the freedom of our will, it certainly encompasses freedom of will as well as freedom of action. Let me briefly explain why we don't control our beliefs and then contrast belief with both judgement and intention.

The intellectualist maintains that we exercise control over ourselves by making normative assessments of our states and activities, both actual and potential. This is, I think, the right conception of control but it does not apply to belief itself (whether normative or non-normative). Why not? Given some initially plausible assumptions about the psychological capacities of a rational person, it seems that we must be able to control our beliefs by forming higher-order beliefs about whether we are entitled to them. If one first assumes that rational belief is based on reasons for belief and then assumes that in so far as belief is sensitive to the reasons for it, it must also be sensitive to our beliefs about those reasons, it seems to follow immediately that, in so far as we are rational, we must be able to control our beliefs by forming beliefs about the reasons for them. Were this so, normative belief would satisfy *Constraint* since normative belief would control itself in just the way that, the intellectualist supposes, it controls action. I agree that rational belief is sensitive to reasons for belief but I deny that that we are

able to control our beliefs by forming beliefs about the reasons for them. Thus normative belief violates *Constraint.*

As already noted, evidence alone does not settle whether someone is justified in believing what they do, and thus whether they know it.[20] One must also consider how much this sort of issue matters to the believer, how confident they need to be on the point. All forms of deliberation, whether doxastic or practical, are subject to constraints of time, energy, and cognitive resource (e.g. memory) and an assessment of the outcome of doxastic deliberation cannot fail to take these limitations into account. Furthermore, the rational believer must himself be sensitive to these limitations. Yet it is a fact that doxastic deliberators lack the capacity to get themselves to form a belief by explicitly considering such factors, by reflecting on whether they should now form a view about whether *p* given these limitations, the importance of the issue, and so forth. At least the deliberator does not have this capacity simply in virtue of being a rational deliberator.[21] So we can't exercise rational control over our beliefs by forming normative beliefs about them (nor indeed by making judgements to the same effect). That is what underlies the widespread idea that belief is not free.

In this respect practical judgment differs from belief, as we can see by returning to our earlier example. There are two kinds of case to consider: those in which we don't take ourselves to know what we ought to do and those in which we do. In both, practical deliberation concludes with a practical judgement, a judgement that can supplement or countermand the operations of belief and one that is itself under the control of practical judgement. This independent capacity for practical judgement is the source of our practical freedom.

First recall my choice of restaurant. Here I don't take myself to know which restaurant is best because I suspect that it could easily turn out that some other restaurant was much better than the one I am presently inclined to choose. I don't even take myself to know whether I ought to make the choice now (though I feel inclined so to do) because if I waited a little longer I just might learn a lot more. Here the rational agent retains the capacity to make a judgement on the basis of what he does know about whether he ought to plump for this restaurant, a judgement which will take account of all the relevant information available to him, including the constraints on the deliberative process itself. He also has the capacity to make a higher order judgement about whether he should make that very judgement, or hold off until he is a bit less distracted for instance. And so on up the potential hierarchy. And, in so far as he is rational, he will act on the judgement with which he terminates the regress even though at no point does he claim to *know* either what he ought to do or what judgment he ought to make.

[20] This paragraph summarizes the argument of Owens 2000: ch. 2.

[21] I am assuming that theoretical deliberation (i.e. the assessment of facts which provide reasons for belief) need not involve beliefs about reasons. See Owens forthcoming.

So, unlike belief (and intention) practical judgement controls itself in just the way it controls action (and intention).

Suppose instead that I have become convinced that a certain course of action is for the best, or at least that it is the one I ought to pursue given the various constraints I labour under. This *becoming convinced* of what I ought to do is quite unlike the making of a practical judgement about what I ought to do. I don't decide to become convinced of this in the way I decide to make a judgement about what I ought to do. Conviction is appropriate where the suspicion that I might easily learn otherwise is inappropriate and what settles this is not just the evidence but also those pragmatic considerations that determine what level of evidence would justify conviction. Yet one can't convince oneself simply by judging, however correctly, that the time has come to make up one's mind, given the constraints one labours under. Rationality doesn't guarantee that you can get yourself into a state where you think you know the answer by means of such reflections. (By contrast, rationality does guarantee that you can get yourself to make a practical judgement on the matter by noting the constraints on your deliberations.)

Suppose I am convinced that such and such is the right course of action. Do I still control my practical judgement on this point or is it now in thrall to this unfree belief? The argument of the previous section preserves my judgemental control over it. It is true of (virtually?) any belief that we have the capacity to judge that we ought not, in this instance, to act as if that belief is true. And we have this capacity simply in virtue of being rational agents. Of course, our judgements are based on an assumed background of default assumptions, that is, on a set of beliefs on which we are relying for present purposes. Nor is this cause for regret: practical freedom would have little value if its exercise were not informed by what we know. But, of each of those beliefs, it is again true that we have the capacity to judge that we ought not to act as if it is true.

I began by saying that, for practical judgement to be the source of our practical freedom, we must have the same sort of control over it that we have over our actions and intentions. I finish by noting that this condition is satisfied. Practical judgement is no more subject to the will than intention. A practical judge can't get himself to judge that *p* solely on the grounds that making that judgement would itself be desirable. But, as we have seen, a rational agent *can* control his practical judgements by reflecting on the constraints on the process of deliberation as well as on the merits of the options. Exactly the same is true of intention. When debating which restaurant to book I won't just be thinking about the relative merits of the restaurants but also about the need to make up my mind sooner rather than later, so that I can lay other plans for the evening on the basis of my choice and so that I can turn my attention to unrelated matters. A rational agent can get himself both to make a judgement about where he ought to eat and also to form an intention to eat there by telling himself that he has thought about the matter for long enough and must now decide

(Owens 2000: 33). He can control both judgement and intention by reflecting on the constraints on the process of practical deliberation. So practical judgement satisfies our requirement: it controls intention (and thus action) in just the sense that it controls itself.[22]

REFERENCES

Arpaly, N. (2003) *Unprincipled Virtue* (Oxford: Oxford University Press).
Bratman, M. (1999) *Faces of Intention* (Cambridge: Cambridge University Press).
Broome, J. (2001) 'Normative Practical Reasoning', *Proceedings of the Aristotelian Society Supplementary Volume,* 75: 175–93.
Fantl, J., and M. McGrath (2002) 'Evidence, Pragmatics and Justification', *Philosophical Review,* 111: 67–94.
——— (2007) 'On Pragmatic Encroachment in Epistemology', *Philosophy and Phenomenological Research,* 75: 558–89.
Gordon, R. (1987) *The Structure of Emotions* (Cambridge: Cambridge University Press).
Greenspan, P. (1988) *Emotions and Reasons* (New York: Routledge).
Hawthorne, J. (2004) *Knowledge and Lotteries* (Oxford: Oxford University Press).
Holton, R. (2006) 'The Act of Choice', *Philosophers' Imprint,* 6: 1–15.
Nagel, T. (1970) *The Possibility of Altruism* (Princeton, NJ: Princeton University Press).
Owens, D. (2000) *Reason Without Freedom* (London: Routledge).
——— (2003) 'Does Belief Have an Aim?', *Philosophical Studies,* 115: 275–97.
——— (forthcoming) 'Deliberation and the First Person', in A. Hatzimoysis (ed.), *Self-Knowledge* (Oxford: Oxford University Press).
Peacocke, C. (1998) 'Conscious Attitudes, Attention and Self-Knowledge', in C. Wright, B. Smith, and C. MacDonald (eds.) *Knowing Our Own Minds* (Oxford: Oxford University Press).
Scanlon, T. (2003) 'Metaphysics and Morals', *Proceedings and Addresses of the American Philosophical Association*, 77: 7–22.
Shah, N., and D. Velleman (2005) 'Doxastic Deliberation', *Philosophical Review,* 114: 497–534.
Williamson, T. (2000) *Knowledge and its Limits* (Oxford: Oxford University Press).

[22] I owe thanks to Matt Soteriou, Nishi Shah, and Richard Holton, to audiences at Amherst College, the Universities of Sheffield, Leeds, and London, and to David Bell for once asking me what I thought a practical judgement was.

7

Two Kinds of Agency

Pamela Hieronymi

I will argue that making a certain assumption allows us to conceptualize more clearly our agency over our minds. The assumption is this: certain attitudes (most uncontroversially, belief and intention) embody their subject's answer to some question or set of questions. I will first explain the assumption and then show that, given the assumption, we should expect to exercise agency over this class of attitudes in (at least) two distinct ways: by answering for ourselves the question they embody and by acting upon them in ways designed to affect them according to our purposes—in roughly the way we exercise agency over most ordinary objects.

The two forms of agency are rarely distinguished, because the first does not display the most familiar and prominent features of agency, while the second might involve an exercise of the first, at two distinct points. Nonetheless, many complex exercises of agency over our minds are easily seen—I think best seen—as composed of these two, more simple, forms. My hope is that decomposing the complex exercises of agency into these two forms might bring some clarity to the difficult topic of mental agency.

1. THE ASSUMPTION

I begin by explaining the assumption. Note that, having settled for oneself some question, one is then in a certain kind of state of mind—namely, a state of mind of having settled that question. For the settling of certain sorts of questions, we give a name to such states. For example, having settled for oneself (positively) the question of whether to ϕ (where ϕ stands for some ordinary action, such as make some lunch or dust the furniture), one therein intends to ϕ.

Note, too, that (for persons, or rational subjects) insofar as one intends to ϕ, one is vulnerable to certain sorts of criticisms and open to certain sorts of questions—in particular, one is open to questions and criticisms that would be satisfied by reasons that (one takes to) bear positively on whether to ϕ.[1] I will

[1] I insert the parenthetical '(one takes to)' because certain of the questions and criticisms (such as Anscombe's famous why-question) would be satisfied simply by whatever one took to settle the

capture this vulnerability with the notion of *commitment*, saying that, insofar as a person intends to ϕ, that person is committed to ϕ-ing. In fact, given that the reasons that would satisfy the questions and criticisms to which one is vulnerable are just those that (one takes to) bear positively on whether to ϕ, it seems that, insofar as one intends to ϕ, one is committed to a positive answer to the question of whether to ϕ.

Thus, if one has settled for oneself positively the question of whether to ϕ, one intends to ϕ, and one intends to ϕ just in case one is committed to a positive answer to the question of whether to ϕ. I will capture this complex conjunction of conditionals by saying that an intention to ϕ *embodies one's answer to the question* of whether to ϕ.

It seems these same claims hold of belief. If one settles for oneself positively the question of whether p (where p stands for a proposition, such as 'The butler did it' or 'All cats are sweet-tempered, deep down'), then one believes p. Likewise, insofar as one believes p, one is committed to a positive answer to the question of whether p, i.e. one is vulnerable to a range of questions and criticisms that would be satisfied by reasons that (one takes to) bear positively on whether p. So we can say that a belief that p embodies a positive answer to the question of whether p.

Far more controversially, I think the same sort of claims can be made about certain emotions—that e.g. one's resentment of S for ϕ-ing embodies one's answer to some range of questions about S's ϕ-ing. I will not defend this more controversial claim, here. Though I think it illuminating—it can shed light both on the nature of certain of our emotions and on the nature or status of the claims I have made about belief and intention—the more controversial claim is not needed, for the main point at hand.

2. TWO KINDS OF AGENCY

If an attitude embodies our answer to a question or set of questions, then it seems we will form or revise such an attitude in forming or revising our answers to the relevant question(s). As noted, if you become convinced that p, and so settle for yourself the question of whether p, you therein, *ipso facto*, believe p. Likewise, if you settle (positively) the question of whether to ϕ, you therein, *ipso facto*, intend to ϕ. Moreover, if you change your mind about whether to ϕ, or about whether p, in such a way that you are no longer committed to ϕ-ing or to

question, while other questions and criticisms (such as certain kinds of moral criticisms) would be satisfied only by reasons that in fact settle the question, while still others (such as certain concerns about justification) would be satisfied by reasons that would settle the question, given your (actual or idealized) epistemic situation. While this complexity is important, for the matter at hand what is crucial is that the questions and criticisms would all be satisfied by considerations that either do bear, would (given certain assumptions) bear, or were taken to bear on a certain question, namely, whether to ϕ.

the truth of p, then you no longer intend to ϕ or believe that p. We might say that we control these aspects of our minds because, as we change our mind, our mind changes—as we form or revise our take on things, we form or revise our attitudes. I call this exercising *evaluative control* over the attitude.[2]

Though this is, I think, the ordinary and most fundamental way of controlling these attitudes, it is far from an ordinary notion of control or agency. In fact, there are a number of important questions one might raise about it and a number of important objections to calling it a form of control. I will examine some of these objections after considering another form of agency we also exercise with respect to these attitudes. Note that these attitudes, which (I claim) embody a person's answer to a question or range of questions, also interact in more-or-less predictable ways with their environment. Our attitudes share this feature with ordinary objects, like chairs, coffee cups, and computers. Insofar as we can think about these attitudes and understand their interaction with their environment, we can control them in the same way we control anything that we can think about that interacts in more-or-less predictable ways with its environment: we can take actions designed to affect them according to our purposes. Our ability to thus control our attitudes is limited only by our cleverness, strength, luck, and industry, i.e. by the same features that limit our control over any object. Thus these attitudes can be objects of a far more familiar sort of control, which I call *managerial* or *manipulative control*.

While it might seem surprising that we can exercise the same form of control over our attitudes that we exercise over more ordinary objects, it should not. Consider the relative ease with which we exercise this familiar form of control over the attitudes of others. If you want to bring it about that someone else believes p or intends to ϕ, you will not, generally, be at a loss as to how to proceed. Of course, in certain cases, for certain values of p or ϕ, the task may be too difficult to achieve. But for a great many values, it will be quite doable—you must simply bring it about that the person settles positively the question of whether p or whether to ϕ, and there is a familiar range of ways to accomplish this. It should not be surprising, then, that we can exercise the same sort of control over our own attitudes—that we can take steps designed to bring it about that we believe p or intend to ϕ. In order to succeed, we have to bring it about that we have answered positively the question of whether p or whether to ϕ—we have to bring it about that we are committed to p as true or to ϕ-ing. In certain cases, for certain values of p or ϕ, this will be quite difficult. For others, it will be relatively obvious what to do.

There is, of course, a kind of difficulty in one's own case that one does not encounter in managing or manipulating the attitudes of others. In order to

[2] This is what Richard Moran sometimes calls 'deliberative' or 'rational' control. See Moran, 2001: esp. 113–20. I do not follow him in using that label, since it seems to me to suggest that this kind of agency requires deliberation or reasons.

bring it about that another person believes or intends, you might provide that person with considerations you predict that person will find compelling, which considerations you do not, yourself, find compelling. But in order to bring it about that you, yourself, believe or intend, and to do so by providing yourself with reasons, you must provide yourself with reasons that you predict you will, yourself, find compelling. But, of course, if you thought there were available compelling reasons, it would be likely that you *already* believe or intend. So the opportunities for managing one's own attitudes by providing oneself with compelling reasons will be more restricted than the opportunities to do so to another. Still, they can arise. If, unable to sleep, you want to believe that your children arrived home safely through the storm, you might call them and so provide yourself with convincing evidence that they have arrived.[3] If you want to be sure that, tomorrow, you will still intend to avoid dessert, you might act, today, to create extra incentives: you might make bets with your friends. Moreover, providing reasons for yourself is not the only way in which you might manage or manipulate a belief or intention. You might undergo hypnosis, or induce amnesia, or convince yourself that an alternative interpretation of your situation is equally justified, and so successfully change your attitudes.

Thus it seems we can manage or manipulate our own attitudes in roughly the way we can manage or manipulate ordinary objects: by taking actions designed to affect them according to our purposes.

3. EVALUATIVE CONTROL AND OBJECTIONS THERETO

I return, now, to evaluative control. I claimed that certain attitudes embody one's answer to a question or set of questions, and that, therefore, one can exercise control or agency over such attitudes by coming to or revising one's answers to the relevant question(s). I acknowledged that it is a far-from-ordinary notion of control. I will here briefly consider a few objections to it, hoping thereby to display its operation more clearly.[4]

[3] I owe this helpful example to Thomas P. Kelly.

[4] There are two possible ways of elaborating upon this view about our agency over our attitudes. On the first, we would distinguish between *settling* a question and *being committed to* an answer to that question. Settling the question, one might think, it is an activity that one may or may not engage in; being committed to an answer to a question is not an activity, but rather some sort of 'normative status'—one is committed just in case one is open to characteristic certain sorts of questions and criticisms (again, questions and criticisms that would be satisfied by reasons (one takes to) support a certain answer to a question). If one has this 'normative status,' it seems we can say that one is committed to an answer to this question. Thus, on this interpretation of evaluative control, we would insist (relatively uncontroversially) that a person is committed to an answer to (a) certain question(s) just in case that person has a certain attitude, but we would allow that one might have the attitude without having engaged in the activity of settling the question. Thus, if one does settle the question(s) for oneself, one exercises control over the attitude. However, someone

Recalcitrance

One might first object that we sometimes settle certain questions without thereby altering either our attitudes or the commitments they entail and are entailed by. Thus, one might think, evaluative control is at best not entirely reliable, and, moreover, I should revise or qualify my claim that, if you settle for yourself this or that question, you therein form this or that attitude. But this claim was, it seems, the main motivation for claiming that we exercise evaluative control.

My reply to this objection will seem, at first, cheap: insofar as you have in fact settled a question, to that extent you do change your commitments (i.e. the questions and criticisms to which you are answerable), and insofar as you have not changed your commitments, you have not in fact settled that question. But, insofar as you have changed your commitments, you have formed or altered the associated attitude. So, if you have in fact settled a question, then you must have formed or altered the associated attitude.

This reply may seem cheap, because it may seem that I am simply defining 'settling a question' so as to ensure my claims are correct, against an obvious, intuitive problem. So I will try to show that, even in the problematic cases, my seeming stipulation is plausible.

By adopting my seeming stipulation, we preclude the possibility of settling a question without therein changing one's commitments and so one's attitudes—but such cases seem not merely possible but actual. We can identify

might have that attitude, and so be committed to an answer to the question—someone might have the 'normative status'—without ever having settled the question, and perhaps without ever having exercised any agency with respect to the attitude. On this interpretation, the attitude embodies one's answer to a question, but it does not, thereby, embody an exercise of agency.

On a second, more radical, interpretation, we would not allow (in persons) being committed to an answer to a question to part company with having settled that question. Rather, we would insist that, if you are committed to an answer—that is, if you are open to those questions and criticisms that would be satisfied by reasons that (you take to) bear on a question or set of questions—then you must have settled that question. On this interpretation, the 'normative status' cannot appear apart from an exercise of agency. Rather, an exercise of agency (viz., the agency at work in settling a question for oneself) incurs the commitment, in each case. (This more radical interpretation would simplify the complex conjunction given above as a definition of 'embody an answer to a question.')

This more radical interpretation will obviously require positing an exercise of agency in a surprisingly wide range of cases, and so, one might think, either will be wildly implausible or else will require an objectionably deflationary account of agency—agency will be attributed wherever we find an attitude with a certain sort of 'normative status,' regardless of whether we find, there or in the agent's history, any discernible mental processes or activities that we could independently identify as an exercise of agency which we might associate with that attitude. Though I am currently inclined to think that we should prefer the more radical interpretation and accept the unusual understanding of agency it entails, defending this (initially implausible or deflationary) choice will require considerable work, and I will not here undertake the task. Rather, I will note that, on either interpretation, one exercises agency over certain attitudes in settling questions for oneself. In the text I concern myself with this weaker claim, which raises enough worries for present discussion. (For an excellent discussion of some of the difficulties that might plague the stronger claim, see Matthew Boyle, 'Making up Your Mind' (manuscript).)

them because we are sometimes able to identify the settling of a question apart from the presence of certain attitudes: you can e.g. settle a question by engaging in a conscious, overt process of deliberation on that question, and coming to a conclusion.[5] But, of course, you might deliberate and come to a conclusion that is at odds with the attitudes you continue to hold. You might believe not *p*, or intend not to ϕ, and then deliberate about whether *p*, or whether to ϕ, and reach a positive conclusion. You might nonetheless continue to believe not *p*, or intend not to ϕ. And this, one might think, shows that you can settle a question without changing your commitments or controlling your attitudes.

I agree with all but the last claim. If you have, in fact, concluded that *p* (e.g), then it seems to me that you will, at least for a moment, incur the commitments associated with believing *p* and, therefore, that you do, at least for a moment, believe *p*—perhaps despite the fact that you also continue to believe not *p*. Thus, in the problematic situation, either you have, upon reaching your conclusion, arrived at the conflicted and difficult state of believing *p* and also believing not *p*, or else you are momentarily waffling in your beliefs about *p*. Saying either seems to me more plausible than saying that you have somehow come to a conclusion without changing your commitments and therefore your attitudes.[6]

Of course, if we accept either of the preferred descriptions, it will be true that, in coming to the conclusion that *p*, you *have* exercised a kind of control over your mind—you formed a belief that *p* in settling for yourself (positively) the question of whether *p*. What you have not done is exercise control over your belief that not *p*. In order to control *that* belief, it seems you will have to find a way to keep yourself consistent—but, importantly, keeping yourself consistent is not required for an exercise of evaluative control.

This response may seem disappointing. To really control your attitudes, one might think, you should be able to target a specific belief—the belief that not *p*, say—and see to it that *that* belief changes, when you settle a question, if you think it should. Short of this, one might think, what I am calling evaluative control does not deserve to be called a kind of control. I will take up this kind of worry next.[7]

[5] It is important that this is not the *only* way that you can settle a question. But it is one way.

[6] Better, it seems, to locate the difficulty in the particular thinking subject, who is conflicted or inconsistent, than to allow that one can settle a question without incurring the associated commitments or to allow that a person's commitments and attitudes can part ways. But someone might disagree about this last claim. I consider such disagreement in the next footnote.

[7] The reply might disappoint in another way: the example may seem to call into question my claim that a person has the attitude just in case that person is rightly open to certain questions and criticisms (i.e. committed). Perhaps, one might think, a person can be (momentarily) subject to criticisms (as a result of settling a question) without therein (momentarily) believing, or have a belief without incurring the typical commitments. I resist this position in part because it seems to me that an attitude, in a person, that does not support the relevant commitments will not be a belief or intention, but rather a thought, fixation, wish, or inclination—something less than a person's belief or intention—and in part because it seems to me that one cannot rightly be subject to the relevant questions and criticisms unless one in fact believes or intends.

The Paradigmatic Features of Agency or Control

So, I hope it relatively plausible that, as you settle for yourself certain questions, you therein, *ipso facto*, form or revise certain attitudes. As you make up your mind about what is true, or what to do, you therein, in some sense literally, make up your mind—you create or constitute, form or revise, your beliefs and intentions. These attitudes, one might say, just are your take on their object, and so, when you change your take on their object, you therein change these attitudes.

While such simple reflections lead naturally to the thought that a thinking subject controls its thoughts (or, at least certain of its thoughts) as it thinks them, there is some reason to resist calling this a form of control, because some of the most salient features of the paradigmatic instances of agency or control are lacking, in this case.

Agency is paradigmatically exercised in ordinary intentional action. *Control* is typically exercised by some subject on some object, where, paradigmatically, the subject has some intentions about the object and controls the object by successfully executing those intentions. Thus it seems that one paradigmatically exercises agency or control by (successfully) executing one's intentions with respect to an action or object. Thus we are led to expect certain features of any exercise of agency or control: we expect exercises of agency or control to display both a certain kind of *voluntariness* (in one sense of that difficult word) and, relatedly, a certain kind of *reflective distance* or *awareness*. But evaluative control displays neither of these features. The forming and revising of beliefs and intentions is not voluntary nor does it require the same kind of reflective distance or awareness.

To illustrate, consider first ordinary intentional actions (such as getting some lunch or managing one's finances). When we intend to do something, it seems we have, in some sense, settled for ourselves positively the question of whether to do that thing—a question that represents the action, under some description. In settling that question, we form an intention, which intention we will, if all goes well, execute in intentional action. Moreover, we can settle the question of whether to ϕ, like any question, for any reason(s) we take to bear convincingly on it—or perhaps for no reason at all. I can e.g. decide to get some lunch for any reason(s) that I take to settle the question of whether to do so. Thus I will say that ordinary actions are *voluntary*, in the following, somewhat technical, sense:

However, perhaps surprisingly, I suspect that granting this objection, and so allowing commitments and attitudes to part ways, would complicate but not entirely upend the view here presented. On the more complicated version of the view, evaluative control would be exercised over the commitments, which would in turn bear some relatively close but not necessary connection to the associated attitudes. Thus, on such a view, one will not only have to keep oneself consistent, somehow, by means other than evaluative control, but also keep one's attitudes in line with one's commitments, somehow. (While this seems a possible view, I would prefer to keep the commitments more clearly associated with some psychology; some such association seems inevitable, and belief seems a good candidate for the job.) I devote myself, in the main text, to what I think is the more pressing and illuminating objection.

we can, for any reason that we take to count sufficiently in favor of the action (or perhaps for no reason at all), settle the question of whether so to act, therein intend so to act, and, providing as all goes well, execute that intention in action.

A certain kind of reflective distance or awareness goes hand-in-hand with this kind of voluntariness: if we form our intentions by settling for ourselves a question that represents our action under some description, then it seems that our action is, in some sense, an object of our thought—in a way that e.g. the unforeseen consequences of our actions are not.[8]

The same features appear in the paradigm cases of control over ordinary objects—over cups and cars and computers. Since we control these objects by forming and successfully executing intentions with respect to them, it seems that the ordinary objects of ordinary control are, in the paradigmatic cases, represented or implicated in the question we settle for ourselves, and, again, we can settle that question for any reason(s) we take to bear convincingly on it.[9]

Thus, in the paradigm cases of agency or control, that over which we exercise control or agency—whether an action or an ordinary object—is, in some sense, a part of the content of our thought, in a way that e.g. the unforeseen consequences of our actions are not. In the paradigm cases, there is a certain familiar reflective distance between the subject who controls and the object that is controlled, or between the agent and what the agent affects (or effects). We exercise agency or control, one might say, when we are the cause of our own representations—the cause of that which we represent as to be done.[10] Moreover, in this 'reflective distance' we encounter a kind of voluntariness: in reflecting upon the action or

[8] It may seem problematic to move from the claim that one is committed to an answer to a question that represents the action to the claim that one represented that action in thought. It seems I have moved from a claim about the criticisms to which one is rightly vulnerable to a claim about what sort of events have occurred in one's mind. While I am tempted to make such moves, I do not think this one is strictly necessary for the point at hand. I think it clear enough that, if we act intentionally, the action we intend is (paradigmatically?) represented to us in a way that unforeseen consequences of our actions are not represented. This will contrast with the attitudes themselves. It seems that our beliefs (i.e. our own states of mind) are not represented, as we form them.

[9] This claim that an ordinary object of control is represented in the question settled will be more controversial than the claim that the intentional action is so represented. After all, it seems you will control your pen in executing an intention to write a note. It does not seem that the pen is represented in the question of whether to write the note. Still, I think it plausible to say that, at some point, your intention to write the note will involve some representation of the pen, since your use of the pen was not unforeseen. Perhaps your intention to write a note leads to an intention whose content has something to do with your pen. There are various ways to understand such a 'nesting' of intentions. For discussion of related issues see Bratman (1987) and Anscombe (1957: 37–47).

[10] Notably, Kant defines the capacity for desire as 'the capacity to be by means of one's representations the cause of the objects of those representations' (*Metaphysics of Morals,* 211, and *Critique of Practical Reason,* v. 10). Many seem to find being the cause of one's representations a necessary, but not sufficient, feature of agency: we are agents, they think, when we not only cause what we have, in some way, represented, but when we do so intentionally—when cause something we have represented because we have in some way decided to cause it. On such a picture, to be an agent is to be able to cause the objects of your representations voluntarily: to be able to exercise a kind of executive capacity over which of your desires is actualized (over which of your representations are the cause of that which they represent).

object of control, we can decide to do that which we have in mind to do for any reason we take to settle the question of whether to do it.[11]

Evaluative control displays neither of these familiar features: the objects of evaluative control (beliefs and intentions) need not stand at a reflective distance in our thought as we exercise this form of control over them—we need not represent our beliefs and intentions in the way we represent our actions. Nor is their formation or revision voluntary in the way ordinary intentional actions are voluntary—we cannot form or revise or maintain or create them for any reason we take to count sufficiently in favor of so doing, but only for reasons proper to them: only for reasons that we take to show the belief true or to bear on whether to perform the action intended.[12] Nonetheless, I think we should grant that we do exercise a kind of control or agency over our attitudes as we settle for ourselves the questions they embody.

I will elaborate on these claims in a moment, while defending the thought that evaluative control should be counted as a kind of agency. But first, to avoid confusion, it will help to note that, on the account presented here, many exercises of mental agency will be instances of what might be called *mental actions*. That is, many exercises of mental agency will share the structure and display the familiar features of ordinary intentional action. So e.g. you might call to mind where you put your keys, try to remember the last time you visited your sister, rotate an object in your imagination, or picture your living room walls a different color. So long as such imaginings and rememberings are intentional, they can be classed, on the account here presented, with ordinary actions like raising your right hand or getting some lunch.[13] No doubt there are many interesting and important questions about the various forms of mental action, but I will not address them here. Rather, I will simply class mental actions with other ordinary actions, and contrast them with the agency at work in the formation and revision of such attitudes as belief and intention.

Doing without the Paradigmatic Features

Why should we allow that what I am calling evaluative control deserves to be thought of as a kind of agency? There is much to be said, but I will confine myself to some brief remarks.

[11] It may be worth noting that the discretion here does not include the ability to do something even when you are convinced that you have sufficient reason not to do it.

[12] Another important dissimilarity: if I am right about the relation between settling a question, incurring a commitment, and having an attitude, then the relation between settling a question and forming the attitude that embodies one's answer is not causal, but rather something like conceptual or constitutive. There is no possibility of things going wrong, between one and the other.

[13] Of course, it may be that you remember something, or that something appears in your imagination, unintentionally—the thought comes unbidden, so to speak. I presume that these mental goings-on need not be treated as instances of mental agency, and so leave them aside.

First, to avoid verbal dispute, it should be granted that one might well reserve the word 'agency' for those activities that do display the familiar features of voluntariness and reflective distance. Such usage would be unobjectionable, so long as it does not invite the thought that anything lacking the distinctive features must be a kind of passivity, or something merely acted upon. Thus, I would insist that some title should be granted to evaluative control (perhaps we could call it a kind of 'activity') that prevents its exercise from being grouped with those things that merely happen to one and prevents its outputs—the attitudes I claim one forms or revises by means of its exercise—from being grouped with those things that one can affect only by acting upon them.

It seems to me plain that we need some additional category of agency or activity—one that does not share the characteristic features—in order to accommodate the agency we exercise over our own intentions. We have already granted that the formation and revision of intention is not voluntary. In fact, I have argued elsewhere that it could not be.[14] Nor, it seems to me, need intentions be represented in thought as one forms or revises them. But if forming and revising an intention does not display the familiar features of the paradigmatic exercises of agency, then it seems that these features could not be essential to agency—since it seems we must exercise some form of agency in forming and revising our intentions, if we exercise agency at all.

One might wonder why I should claim that intention is not, and could not be, voluntary in the sense at issue. I will first present a pair of cases that I hope will lend the claim some intuitive support and then briefly sketch the argument I have given elsewhere.

Consider, first, a case that seems to suggest that intention *is* voluntary. Suppose an experimental psychologist with an 'intention-detector' offers you a small sum for intending to drink some water. It seems you can decide to form the intention and earn the money. Thus it seems that intending is like raising your right hand—something you can do on command, as a so-called 'basic action.' More to the point, it may seem that intending is voluntary in the way ordinary action is voluntary: it may seem that you can decide to intend for any reason that you think shows intending worth doing.

Now suppose instead that the psychologist would like to see register, on her machine, an intention to jump from the third-storey window, and she offers you the same small sum for forming that intention. You might well think the small sum is well worth *intending* to jump (no harm, you think, in simply intending). But, of course, you will not intend to jump, and so will not earn her reward, unless you are committed to *jumping*—unless you have settled for yourself positively the question of whether to jump. And the small sum is not, you think, reason enough to settle that question. But if you do not think the sum reason enough to settle the question of whether to jump, then (assuming

[14] See e.g. Hieronymi (2006, 2008).

that you have no other reasons for jumping and some reasons not to jump) it seems that you cannot respond to her offer by intending to jump, and so cannot earn her money. So it seems you cannot, in this case, intend for reasons that you think count sufficiently in favor of intending.[15]

What prevents you from earning the reward, in the second case? I suggest it is the fact that one intends only if one is committed to an action, but one cannot become committed to an action by finding convincing reasons that one only takes to show the intention good to have. That is to say, one cannot become vulnerable to questions and criticisms that would be satisfied by reasons that (one takes to) bear on whether to ϕ by finding convincing the reasons that one does *not* take to settle this question, but which one rather takes to settle the distinct question of whether the intention to ϕ is good to have. Thus, one cannot form an intention for any reason that one takes to count sufficiently in favor of intending; one can only form an intention for reasons one takes to settle the question of whether to act. In contrast, one can act for any reason one takes to count sufficiently in favor of acting. Thus, intending is not voluntary in the way ordinary action is.[16]

Why, then, can you earn the reward in the first case? We can give the following interpretation. When the psychologist offers you the small sum to intend to drink the water, you can take the offer to be reason enough to settle the question

[15] The case is science-fictional, but it need not be. There are plenty of everyday cases in which a reason for an intention is not reason enough to act. Perhaps it displeases you that I do not intend to finish my work by tomorrow. And perhaps you would be satisfied simply knowing I intend, regardless of whether I actually finish. And perhaps I am generally happy to house mental states that please you. Still, I will not be able to intend to finish, in order to please you, unless I also take pleasing you to be reason enough, not just to house the intention, but to finish.

These cases are, of course, variations on Kavka's Toxin Puzzle (Kavka 1983). The case of intending to jump out the window is unlike Kavka's puzzle, in that, in Kavka's puzzle, the reward for the intention is well worth performing the action (in this way Kavka's case is like the case of drinking the water). The case of drinking the water is unlike Kakva's puzzle in that the action is performed immediately and carries no disincentive. I consider Kavka's puzzle in a lengthy footnote in Hieronymi (2006).

Niko Kolodny points out that, in any such example (science-fictional or no), any reason against acting will also be a reason against intending so to act, since your intentions are likely to lead to action. Thus, he thinks I have not yet provided a case in which you have sufficient reason to intend though you lack sufficient reason to act, and so he remains unconvinced of my claim that you cannot intend for any reason that you take to count sufficiently in favor of intending. For all I have said, it may still be the case that you can intend for any reason you take to count sufficiently in favor of so doing. I grant that the examples do not establish the claim. For further treatment of Kolodny's objection, see n. 17.

[16] For reasons that one takes instead to show an intention good to have (which one does not take also to show the action worth doing), one *could* form an intention *to bring it about* that one forms the desired intention—one could, by finding convincing reasons that one takes to show an intention good to have, commit to the action of bringing that intention about. But, again, one need not make such managerial commitments in the case of ordinary action: ordinarily one need not form an intention to bring it about that one acts; one simply forms an intention to act, and executes that intention in the action. Thus, again, intention is not voluntary in the way that ordinary action is. The argument of this paragraph appears in both Hieronymi 2006 and 2008.

of whether to drink, therein decide to drink, and so intend and earn the money. But you cannot do the same, in the second case, because you do not think the small sum is worth the jump.

If this treatment of these cases is correct, then intending is not voluntary in the way an action is: you cannot form, revise, or maintain an intention for any reason you think counts sufficiently in favor of forming, revising, or maintaining it. Rather, you can only form, revise, or maintain an intention for reasons that you take to settle the question of whether to act. But you can act for any reason you take to count sufficiently in favor of acting. And so it seems that intention is not voluntary in the way that ordinary action is.[17]

[17] Sometimes, in response to this sort of argument, people insist that you can form, revise, or maintain an intention for any reason you take count sufficiently in favor of doing so, but add that the question of whether to form, revise, or maintain an intention to ϕ is 'transparent to' the question of whether to ϕ—that these questions must be answered by the same set of reasons. In this case, asking yourself whether to intend to ϕ seems simply to be a (somewhat sophisticated, reflective) way of asking yourself whether to ϕ. It is sophisticated or reflective (at least) in that it brings to one's mind the fact that, if one decides to ϕ, one will, therein, intend to ϕ. (This is closely related to Kolodny's objection, above.)

(Richard Moran (2001) developed an account of transparency in his investigation of self-knowledge. Notably, Moran thinks the question of whether I believe p (e.g.) is, insofar as I am rational, transparent to the question of whether p—i.e. these questions will be settled by the same reasons. Nishi Shah (2003) considers a transparency thesis closer to the one here considered.)

Notice, though, that if we insist that the question of whether to intend to ϕ can be settled only by reasons that bear on whether to ϕ, it seems that we have given up the thought that intending is voluntary in the way that ordinary action is voluntary. An ordinary action is voluntary in that it can be done for *any* reason that one takes to count sufficiently in favor of so acting. But, on the interpretation just given, intending to ϕ cannot be done for any reason one takes to count sufficiently in favor of so intending. Rather, one can decide to intend to ϕ only in those cases in which one can decide to ϕ.

One might, at this point, return with Kolodny's objection. Recall that Kolodny doubted that there would be cases in which one has sufficient reason to intend to ϕ but lacks sufficient reason to ϕ, because ϕ-ing is a(n obvious) consequence of intending to ϕ. So, the bad effects of jumping show that you do not have sufficient reason to intend to jump. Following this line of reasoning, one might think that intending might be voluntary after all: maybe you *can* intend to ϕ for any reason that counts sufficiently in favor of so doing. It just turns out that you will have such reasons only in cases in which you also have sufficient reason to ϕ.

But even if one established that the only considerations that in fact count sufficiently in favor of intending are those that count sufficiently in favor of acting, and so established that a person can intend to ϕ for any reason that (in fact) counts sufficiently in favor of so doing, one would not thereby undermine my claim. My claim is that, while you can (intend to act, and, providing all goes well) act for any reason that you take to count sufficiently in favor of so acting, you cannot intend to ϕ for any reason that *you take to* count sufficiently in favor of doing intending. So, to undermine my claim, one would have to establish, not just that the only reasons for intending are those that are (in fact) reasons for acting, but that no one could *take* reasons to count sufficiently in favor of intending without also *taking* them to count sufficiently in favor of jumping. (Shah (2003) is aiming at something like this position, with respect to belief). But it seems possible that someone might take that view, even if it is mistaken. So, suppose someone (perhaps mistakenly) thought that the small sum counts sufficiently in favor of intending to jump, without taking it to count sufficiently in favor of jumping. My claim is that such a person cannot intend for the reasons that she takes to count sufficiently in favor of intending, though she could (providing all goes well) jump for *any* reason that she takes to count sufficiently in favor of jumping.

To put the point another way: you will intend to ϕ only if you are committed to ϕ-ing, and (if you commit to ϕ-ing for reasons) you can only commit to ϕ-ing for reasons that you take to settle

Once we grant, however, that forming an intention is not voluntary, it seems that we cannot require that every exercise of agency be voluntary: because it seems that the forming of an intention must be an exercise of agency, if anything is.

One might grant that voluntariness of the sort specified is not essential to exercises of agency, but hold out for the other familiar feature: the reflective distance or awareness. Though many seem to find this feature important, it seems to me inessential. I will briefly suggest why.

Consider, again, intention. Insisting that an exercise of agency must involve the characteristic reflective distance or awareness requires us to say that the forming of an intention was not an exercise of agency unless the agent had some thought about or awareness of that intention—indeed, unless the agent had a thought about or awareness of the intention *of the sort* characteristic of our thought about or awareness of our own actions or the ordinary objects we control thereby.[18] But it seems to me implausible to claim that we are typically, or even very frequently, thus aware of or reflective about our own minds (as opposed to the actions we intend or the ordinary objects we control). Further, it seems that a lack of such awareness of our minds does not distract from the agency we exercise in acting.

Suppose, to be fanciful, that someone is part of a psychological study in which she is taking a drug that will make her nauseous if she forms an intention to stay up late. She is now, under stress, trying to figure out how to finish all the projects she must accomplish by the end of the week. In trying to work out this practical problem, she plans to stay up late tonight, but she does so while forgetting not only that forming such an intention will make her nauseous, but also unmindful, even, of the fact that she has just formed an intention—unmindful of the fact that she has just changed her psychology. It seems to me that, in this case, not only the bad effect but even her intention itself is an unforeseen consequence of her attempt to solve her practical problem. And yet, for this lack of awareness of her own mind, her decision to stay up late seems no less an exercise of her agency.[19] And thus it seems to me that we can

the question of whether to ϕ. But you might (perhaps mistakenly) take certain considerations to show intending to ϕ worth doing, which you do not take to show ϕ-ing worth doing. You will not be able to intend for these reasons (though, as noted, you may be able to bring it about that you intend for those reasons). In contrast, you can (intend to ϕ and, providing all goes well) ϕ for any reason you take to show ϕ-ing worth doing.

[18] Some will want to insist that an intention occurs in its own content, and so think that they have secured for intention the paradigmatic feature of ordinary action. While there may be other reasons for insisting that an intention occurs in its own content, I doubt that this strategy can plausibly gain for intention the sort of awareness that is characteristic of ordinary action or control over ordinary objects.

A full treatment of this claim will obviously require some account of how we are aware of our actions and the objects we thereby control. I have given some indication of my account of this, above: we settle a question that represents our action, under some description.

[19] I am very grateful to Yannig Luthra for his thoughts on this example, and for pressing for clarification.

exercise agency with respect to our intentions even when we are not aware of them in the characteristic way in which we are typically aware of our intentional actions.[20]

Allowing that we can be agents with respect to our attitudes even when we are not aware of or reflective about them goes against a powerful intuition. It seems very odd to think that we can be exercising our agency—and do so normally and well—by creating something that we did not intend to create and that remains, so to speak, out of our own view, behind our back, or off-stage—something that may well be 'unforeseen.' But once we notice that, whenever we make a decision or come to a conclusion on some topic, we therein make something true of our own minds (namely, that we have decided or concluded); that we can do so without having any intentions about our own minds; and that, even if we were to turn our attention to our own minds and to make decisions and come to conclusions or form intentions about it, we would, in so doing, create a higher-order set of attitudes with the same disquieting features, we might start to think that this intuition is simply a bias born of our familiarity with our agency as exercised in our actions and over ordinary objects. I believe we should go without it.

I hope, then, that I have at least suggested why we might allow that evaluative control is a form of control or agency, despite the fact that it lacks the familiar features. I will now, as promised, consider how managerial control can seem to involve an exercise of evaluative control, at two distinct points, and how certain familiar, complex exercises of agency over our own minds can be more clearly understood as so composed.

[20] So-called 'Freudian slips' provide a different kind of example in which one intends without awareness of one's intention; these are sometimes taken to show that 'full-blooded' agency requires some awareness of one's own intentions and/or motivations (see e.g. Velleman 1993). But it should be noted, I think, that in such cases one is (also) unaware that one is doing the action in question (either at all or, at least, under the description under which the action is a 'slip'). So, even if it is granted that these are not cases of full-blooded agency, this might show only that full-blooded agency, when exercised in action, requires an awareness of the description under which one is in fact acting, not that it requires an awareness of one's state of mind, or of one's motivations, or an awareness that, as one decides to act, one is making certain things true of one's psychology.

We should wonder why awareness of one's intention would be thought to make one's agency over one's action more full. There is, of course, one way in which such awareness enhances one's agency over one's action: if one is aware of the fact that, in deciding to act, one will therein change one's state of mind, then one is more fully aware of both the possible reasons for and possible consequences of one's action. Being so aware, one can e.g. decide to drink in order to form an intention and earn the small sum, or decide against ϕ-ing in order to avoid the bad effects of an intention to ϕ. But this is just to say that an awareness of one's own mind can enhance one's agency in acting in just the way that any further relevant information can: I am, in this sense, more fully an agent anytime I am more fully aware of all my options, or all my possibilities—and more fully an agent the less that remains unforeseen. I would readily grant that one's agency in the case at hand is less than full, in this sense. But this can be granted without damage to the point: one exercises agency of an ordinary, non-defective sort over one's action, even when one does not have in mind one's mind.

4. MANAGERIAL CONTROL AND ITS DECOMPOSITION

Consider, first, the great variety of methods by which one might manage or manipulate one's own beliefs or intentions—the variety of methods by which one might take action so as to affect one's beliefs and intentions according to one's purposes.

Most bluntly, you might bring it about that you believe p or intend to ϕ by doing something that affects your brain in a way that is likely to have this effect. If e.g. you want to believe that your friend has never betrayed you, you might induce in yourself amnesia about the relevant stretch of shared history. If you want to believe that this or that is not so worrisome, you might take some anti-anxiety medication. Perhaps, at some point in the future, we will be able to induce particular beliefs or intentions directly, by taking a pill or stimulating the brain. Perhaps hypnosis produces a similar effect.

At the opposite extreme, you might bring it about that you believe p simply by changing the world so as to make p obviously true. As pointed out by Richard Feldman (2000), if you want to believe the lights are on in your office, you can get up and throw the switch.

Less radically, you can also manage your own attitudes by taking steps that you can predict will provide you with convincing reasons for the answer embodied in the attitude. So, again, if you want to believe that your children arrived home safely through the storm, you might call them and thereby provide yourself with convincing evidence that they have. If you want to be sure that, tomorrow, you will do the right thing, you might tell your friends about your plans, today.

(There will be some difficulty, of course, if you believe that you have provided yourself with skewed or unfair evidence for p—because this belief will make the evidence less compelling in your own eyes, and so make it less likely that you will conclude that p on the basis of it, and so make it less likely that you will successfully bring it about that you believe p. As noted earlier, while you can bring it about that someone else believes by providing that person with reasons that you do not, yourself, find convincing, you cannot do the same to yourself. Self-deception is notoriously harder than deceiving others.)

Somewhat more subtly, you might manage your own attitudes, not by providing yourself with new reasons, but by convincing or persuading yourself that the reasons at hand support an alternative conclusion. You might intentionally direct your attention in certain ways, or provide yourself with alternative interpretations of your situation, or persuade yourself to 'see things differently,' or take steps to convince yourself that your own previous response is unjustified or that an alternative response is equally justified. Or you might take steps to keep your attention focused on the reasons you already find convincing, which you predict you will be tempted to overlook in the future: if

you want to strengthen your dieting resolve, you might post a picture on the refrigerator door.

In any of these ways, then, you might take steps to bring it about that you believe p or intend to ϕ, in much the same way that you might take steps to bring it about that someone else believes (or, for that matter, in the same sort of way that you might bring it about that your living room walls are pale green): you take action designed to bring about that end, subject to the ordinary sorts of limitations one always encounters in trying to effect changes in the world.

These cases in hand, note that a successful exercise of manipulative or managerial control over one's attitudes will involve, at two distinct points, a commitment to an answer to a question, and so, it seems, might involve two distinct exercises of evaluative control.

First, if we assume that an exercise of managerial or manipulative control is intentional (as it seems it must be, to earn the title[21]), then, when one exercises managerial or manipulative control over one's attitudes, one will intend so to manage or manipulate one's attitudes, and so will be committed to a positive answer to the question of whether to do so. This commitment would seem to be the result of having settled for oneself the question of whether to manage or manipulate—the result, that is, of an exercise of evaluative control.

Second, if you succeed in your exercise of managerial or manipulative control, you will bring about an attitude that embodies your answer to a question. Thus, if you succeed, you will have brought it about that you are committed to whatever answer is embodied in the attitude.[22] While there is room for disagreement about whether, in bringing about this commitment, you will have brought about an exercise of agency, it should be granted that, at least in certain cases, one can bring it about that someone (perhaps oneself) believes or

[21] This should not confuse: an exercise of managerial or manipulative control must be intentional, to qualify as control, despite the fact that an exercise of evaluative control need not be. Managerial or manipulative control is a matter of acting so as to affect something according to one's purposes. If one acts so as to affect something according to one's purposes without intending to, it seems wrong to say that one has exercised control over that thing.

[22] Of course, one might bring about this second commitment—the commitment embodied in the target attitude—either honestly, so to speak, or dishonestly. This accounts for the continued use of the cumbersome disjunction, 'managerial or manipulative control.' As we saw, you might bring yourself to believe p by making p obviously true, conducting a fair investigation, or providing for yourself evidence, or you might take steps that produce incentives that ensure that you will intend to ϕ, or persuade yourself to take up another, equally reasonable, point of view on x. If you bring about the commitment by any of these 'honest' means, it will seem right to say you *managed* your attitude, or that you exercised *managerial control* over it. But we are not restricted to honest effort. You can bring it about that your attitudes change in ways that produce irrationality or require some kind of self-deception or amnesia. In such cases it will seem right to say that you have *manipulated* your attitude, or that you have exercised *manipulative control* over it. I suspect the distinction between management and manipulation will be hard to draw sharply; happily, we need not draw it sharply, for present purposes. We can simply note that all of these methods belong to a genus: they are ways of acting so as to bring it about that you form or revise or maintain some attitude, which attitude itself embodies your answer to a question.

intends by bringing it about that that person exercises his or her agency in a certain way.

So it seems that any successful exercise of managerial or manipulative control will require a commitment to an answer to a question at two distinct points: one will be committed to a positive answer to the question of whether so to manage or manipulate one's attitude and one will be committed to whatever answer is embodied in the attitude successfully managed or manipulated.[23] Either of these might involve an exercise of evaluative control.

Thus it seems that these two forms of agency can display a characteristic division of labor in an exercise of managerial or manipulative control: perhaps you decide to manage something—in the cases at hand, some attitude of yours, which we will call the target attitude. That decision itself constitutes an exercise of evaluative control with respect to an intention—in deciding to manage the target attitude, you form an intention about it. That intention is executed in a managerial or manipulative action, aimed at changing the target attitude. (Of course, the action here may be mental: it may consist of directing your attention in certain ways, calling to mind certain facts, presenting yourself with an alternative interpretation, or vividly imagining certain outcomes.) Insofar as your managerial or manipulative actions succeed in their aim, you will bring it about that you are committed to the answer(s) embodied in the target attitude. This might involve bringing it about that you settle the relevant question(s) in the relevant ways, and so might involve inducing or influencing the conclusion you come to on some question—that is, it might involve inducing or influencing an exercise of evaluative control.

If we allow that these two forms of control can thus work in tandem, it seems that an exercise of evaluative control can be induced or produced by an exercise of managerial control and that an exercise of evaluative control can initiate each exercise of managerial or manipulative control. Some will find both these claims unsettling or disorienting. The first will seem unsettling because it can seem that exercises of agency should not be the sort of thing that can be induced or brought about or manipulated. But to so insist is to deny not only some of the most important forms of self-management but also some of the most obvious forms of moral wrongdoing (those that involve the manipulation of another's will).[24] The second claim will seem unsettling because it will seem that whatever initiates an exercise of control should display the familiar features of agency: it should be voluntary and should involve reflective distance or awareness. Though I have already suggested why I think we need to allow a form of agency that does not

[23] Or, of course, if one successfully rids oneself of an attitude, one will then cease to be committed to the answer it embodies.

[24] In his presidential address, Rogers Albritton (2003) might seem to suggest that your will cannot be manipulated. But I think he is in fact making a different point: that you cannot, as a conceptual matter, be made to will something against your will—if you are made to will it, you then will have willed it. While this is doubtlessly true, it is hardly a defense against manipulation.

display these features, I suspect that my reflections may not unseat the strong intuition. I will close, then, by briefly considering one popular alternative model of our agency over our minds, one that preserves the familiar features.

5. REFLECTIVE CONTROL

Many philosophers are drawn to a class of accounts of our agency over our minds that I will group together under the head *reflective control*. On such accounts, we exercise agency over attitudes like belief and intention by reflecting critically upon them and determining for ourselves whether they are justified.[25]

Reflective control is attractive, at least in large part, because it seems to preserve the paradigmatic features of ordinary agency. After explaining how it does so, I will suggest that reflective control is difficult to model—it is difficult to understand just how it works. I will briefly mention some ways in which it has been modeled and suggest why I find them dissatisfying before suggesting that we might be able to construct (what seems to me) a more satisfying account by employing the accounts of evaluative and managerial control I offered above (together with an assumption that a proponent of reflective control must also employ). Of course, the account I have offered abandons the thought that our agency over our minds must display the paradigmatic features of agency. Thus, if we use it to model reflective control, it seems we might give up the thought that reflective control is the primary way in which a rational agent exercises her agency over her own mind.

To begin, we need a clearer understanding of the phenomena I am calling reflective control. Many have been powerfully struck by the fact that we can change our own attitudes simply by reflecting on whether they are justified. It is, indeed, a striking fact. After all, reflecting on the justification of this or that does not typically alter the object of the reflection. I may e.g. reflect on some belief or intention of yours, and come to the conclusion that your belief is unjustified or your intention unsound, without thereby having the least effect upon your belief or intention. Moreover, I may even communicate my reflections to you, without thereby changing your attitudes—and neither of us need be, for the ineffective exchange, in any way irrational. We may simply, reasonably, disagree. But if you find, upon reflection, that one of your own beliefs is unjustified or one of your intentions is unsound, then, often enough, that reflection itself seems sufficient to undermine the attitude. Of course, it does not always do so—sometimes you can find yourself in the inconsistent position of believing something you also, in

[25] A nice recent discussion of various proponents of reflective control can be found in Owens (2000). A slightly different set of accounts, also deserving of the name, have it that we exercise agency over our attitudes by determining for ourselves whether we want to have them, or whether they make sense to us. See e.g. the papers collected in Frankfurt (1988) and Velleman (2000).

reflection, have determined unjustified, or intending to do something you also think a bad idea. But in such a case you are, it is said, in some way irrational. Thus it seems, insofar as you are rational, reflecting upon whether your attitudes are justified will, itself, change them.

Note how reflective control seems to share the paradigmatic features of ordinary agency: we can intentionally, for any reason we see fit, decide to reflect upon whether our attitudes are well grounded. Reflecting upon one's own attitudes can be voluntary, and can be done for a purpose. Often enough it is done for the purpose of ensuring that one's attitudes are justified. (Insofar as it is done for that purpose, then, insofar as one is rational, it will achieve its end.[26]) Further, when we reflect on our attitudes, we certainly stand at some reflective distance from them. When we determine whether some attitude of our own is justified, we come to a conclusion about *it*—about the object of our reflection. Still, even though reflective control displays the paradigmatic features of agency, changing your attitudes by reflecting upon their justification seems quite unlike acting upon them in a merely managerial or manipulative way—quite unlike acting upon them in the way we act upon ordinary objects. These facts make attractive the thought that it is reflective control, rather than evaluative control, that provides the best model of our distinctive agency over our own attitudes: we are agents over these attitudes (or perhaps most fully agents over them), not when we simply reflect on and come to conclusions about their content, but when we reflect on and come to conclusions about whether they are justified.

Before adopting this model, I think we need to better understand just how reflective control works—just how is it that reflecting upon the justification of one of your attitudes can change the attitude? Understanding this proves more difficult than is sometimes noted.

Sometimes people talk about the 'authority of reflection' or 'command of reason', as though a reflective judgment serves as an authoritative decree that one's attitudes obey, insofar as one is rational. But such talk is surely metaphorical. Retreating from the metaphor, people sometimes simply say that, when we reflect and find that some attitude of ours is unjustified, we then 'correct' or 'revise' or 'update' the attitude under reflection. But the question at issue is, just how do we accomplish this correction or revision? What sort of activity or agency is exercised in such correction or revision? As already noted, the correction would not be well modeled as an action of the ordinary sort—we do not find ourselves with a bad attitude and then decide to change it by performing some action, as though we were changing a bad spark-plug. Exercising reflective control over one's own mind is not like surveying and tinkering under one's own hood.[27] It is

[26] Absent the stipulation of rationality, reflective control also seems to involve the familiar possibility of failure. With that stipulation, it displays the kind of invincibility sometimes thought to be distinctive of autonomous agency.

[27] This point was suggested to me long ago by Richard Moran.

not, to drop the metaphor, an exercise of managerial control—of taking action so as to affect one's mind.

The difficulty here should not be underestimated: the problem with modeling the correction or revision of one's attitudes as an action is not that it is hard to see what sort of *process* the corrective or revisionary action would involve; the problem is not alleviated by e.g. thinking that the action of correcting one's own attitude is a basic one, which can be accomplished simply by deciding to do it. (Attempting to alleviate the problem in this way will return one to the metaphor of command: one will think that correcting one's attitudes is, after all, like surveying and tinkering under your own hood, so long as you are endowed with godlike powers to effect the required changes just by deciding that it be so.) Rather, the problem is that the correction or revision of an attitude is not well modeled as an intentional action at all, whether basic or complex. First, correcting or revising your own attitudes is not voluntary (in contrast e.g. to correcting or revising your own speech): you cannot decide whether to correct or revise for any reason you think shows it good to do or not do. Perhaps e.g. you think there is very good reason to leave some error in place. It does not seem open to you to do so, given the stipulation of rationality. Further, if the correction or revision were an intentional action, it would be accomplished by settling for oneself the question of whether to correct or revise. The correction, then, seems to be initiated by an exercise of evaluative control—but we were appealing to reflective control precisely to try to understand the most fundamental exercise of mental agency.

A somewhat more promising route employs the thought that, if you find that an attitude of yours—a belief that *p*, say—is unjustified, you will therein form a second-order attitude about that belief: a belief that your belief that *p* is unjustified. (I would say that you form this second-order attitude by an exercise of evaluative control.) One might think that, given this higher-order thought, simple compliance with the requirements of rationality will ensure that you do not go on believing that *p*. Insofar as one is rational, we might say, the lower-order attitude is 'sensitive to' the higher-order judgment.[28] Thus, it might seem, once a rational creature is capable of reflection, it gains a kind of control over its own mind: insofar as it is rational, its mind will conform to its own reflective thoughts about how it should be.[29]

This is a powerfully attractive account. Again, I believe it is powerfully attractive in large part because it preserves the familiar features of the paradigmatic exercises of agency. It is easy to imagine that the agent is the one reflecting, making the higher-order judgment, and it is easy to think of that higher-order judgment as

[28] A powerful presentation of this thought, using the notion of 'judgment-sensitive attitudes' can be found in Scanlon (1998: esp. ch. 1).

[29] The ability to have the mind one wants is the cornerstone of what has been a very fruitful line of thought over the last several decades. See e.g. Frankfurt (1971) and Taylor (1976).

in some way affecting the lower-order attitude, so long as one is rational. Making a judgment and thereby effectively changing an attitude seems a lot like acting upon an object or issuing an effective command, either of which are obvious exercises of agency.

Though this is a powerfully attractive account, it should be noted that the question at hand—just how *does* one correct or revise one's attitudes under reflection—is not clearly answered by it. The account simply stipulates that one has satisfied the standards of rationality, and notes that these standards require a change in attitude. But we were wanting to understand how it is that one changes that attitude, given that doing so is not a matter of performing a kind of mental action. So, even if we grant (what, below, I will suggest we should not) that, if one is rational, one is sure to change one's first-order attitudes upon making the higher-order judgment, we will not thereby have come to understand the agency by which we conform to that requirement. This lack is made more worrisome by the simple fact that one's mind is functioning in accordance with certain standards does not typically show that one has exercised agency. (The well-functioning of one's perception or one's memory e.g. does not seem to be, itself, an exercise of agency.[30])

One might reply that the functioning of a mind in accordance with the standards of rationality *just is* the activity of an agent. While I have some sympathy with this thought, I worry about its implications for our moments of irrationality. Still, even if we were to grant this thought, we might raise other worries about the picture at hand:

First, in order to preserve the familiar features of ordinary agency, rationality must in some way privilege the higher-order, reflective thought over the lower-order, unreflective one. But it is not clear why the requirements of rationality should display such a bias. Perhaps it is your lower-order thought that is rational and reasonable, and your higher-order one that is paranoid, compulsive, or self-deceived. In such a case it seems that the requirements of rationality might ask you to persist in your lower-order thought and abandon the higher-order one.

Additionally, and familiarly, it is unclear why sensitivity to a higher-order thought should render a lower-order attitude the product of one's agency or control unless the higher-order thought is itself already an instance, embodiment, or product of agency, or unless the agent is in some way already identified with it,

[30] One might argue that the revision of the lower-order attitude is an exercise of agency in the following way. The well functioning of one's mind, one might say, ensures that that exercise of agency has its natural effects, in much the way that the well functioning of one's musculature ensures that one's intentions have their natural effects. Thus, just as actions are exercises of agency, so is the revision of the attitude. This line of thought overlooks an important difference between the action and the revision of the first-order attitude, a difference which appears when things do not function well. If one's musculature fails in some way—seizes or spasms—one's action fails, but one is in no way irrational. But if one fails to revise one's lower-order attitudes, one is irrational.

such that its effects can be identified as hers.[31] Taking the first route—explaining why the higher-order thought is itself already an instance, embodiment, or product of agency—will, I think, lead one back to notion of evaluative control (or something very much like it).[32] The second route—identifying the agent with the higher-order attitude, such that its effects are hers—has been taken by a number of people.[33] Notably, those who take this route appeal, not to reflective judgments, but rather to desires or values. Moreover, I believe they do so advisedly: values and certain desires can plausibly be claimed to be the sort of thing with which an agent is essentially identified—something an agent cannot coherently disavow. A higher-order judgment about the justification of some other attitude, in contrast, does not seem to be the sort of thing with which an agent is essentially identified; it seems, rather, like something one can coherently disavow.[34]

Rather than trying to further explain and defend my dissatisfaction with existing accounts of reflective control, I will, at this point, simply present the beginnings of an alternative. I believe the account of evaluative and managerial control that I have offered above might provide a particularly promising way to start to understand reflective control. Insofar as this alternative is plausible, it raises another worry for what I have called the powerful picture.

Consider, then, successfully revising one's belief that *p* under reflection. One believes *p*, we have said, just in case one is committed to a positive answer to the question of whether *p*. To conclude that one's belief that *p* is unjustified is to conclude that one does not have sufficient reason to settle that question positively. Thus, to conclude that one's belief that *p* is unjustified is to settle negatively the question of *whether the reasons available to me show that p*. But it seems that settling this question might involve reconsidering the simpler question, *whether p*, while employing the reasons available to you. That is to say, reaching the conclusion that your belief that *p* is unjustified might itself involve reconsidering the basic question, whether *p*, and failing to settle it positively. If it does, then, insofar as you remain consistent, or of one mind, on the root question of whether *p*, you will, in failing to settle the question positively, therein suspend your belief that *p*. That is to say, insofar as you remain of one mind on

[31] This is structurally similar to the point made by Gary Watson against Frankfurt's hierarchical account of free action. See Watson (1982). Frankfurt and his followers retrenched by attempting to identify the agent with some (set of) attitude(s), which then relate to others. Frankfurt's later work is found in Frankfurt (1988, 1999). J. David Velleman develops the thought in a different way. See Velleman (1989, 2000).

[32] Recall that the appeal to reflection was an attempt to preserve the standard features. We have now found, in attempting to work out the reflective account, that we need to appeal to an exercise of agency within it—agency in forming a judgment. If we try to secure the standard features here, we risk generating a regress.

[33] Central examples are, again, Velleman and Frankfurt.

[34] My treatment of such alternatives here is obviously only provisional. I hope to give them a fuller hearing in later work.

the question of whether *p*, you might revise your belief *in the process* of finding it unjustified.[35]

Understanding reflective control along these lines has several benefits. We have accounted for the change in the first-order attitude by appeal to an exercise of evaluative control together, not with the requirements of rationality (quite generally), but rather with a (weaker) requirement of consistency. We stipulated, not that rationality privileges the higher-order judgment over the lower-order attitude, but simply that the person stays of one opinion on the root question, as he or she settles the more sophisticated question. Further, because it is clear that evaluative control is being exercised as the person addresses the sophisticated question, there is no need to identify the agent especially with the higher-order judgment. We rather simply identify the agent as the one exercising evaluative control.

The proposed account also goes somewhat further than the alternatives in answering the question with which we started: just what form of agency is exercised over the attitudes revised under reflection? The proposed account would have it that one exercises evaluative control in revising one's attitudes under reflection.

Note that, insofar as the alternative picture is close to correct, the original models of reflective control are not just metaphorical, but actually misleading. We started by appeal to the metaphor of commanding or tinkering. But notice that any commanding of or tinkering with attitudes must be *subsequent to* the judgment that the attitude is unjustified: one first makes the judgment and

[35] Yannig Luthra and Sheldon R. Smith independently suggested that a person might conclude that she does not have sufficient reason to believe *p* without re-posing the question of whether *p*: perhaps she now decides that any belief she acquired last night, when exhausted and under the influence of all those pain-killers, must be unjustified. And suppose she knows that last night she acquired the belief that she will recover fully from her accident in two weeks time. She might now conclude that she does not have sufficient reason to believe that she will recover fully in two weeks time, and thereby lose that belief. But it might seem that she has revised her belief without re-posing for herself the question of whether she will recover fully in two weeks—she answered for herself a question about the justification of her belief, without reconsidering the truth of the matter (as Yannig nicely put it, she has reasoned as a juror, about the adequacy of her evidence, not a detective, about the facts)—and so this might seem a different kind of case than the one I consider in the text. If so, then in such a case the question I have been asking remains unanswered: how, exactly, does the person revise her unjustified belief? What kind of agency is at work, in the revision?

But I am not sure that, in drawing the conclusion that, because of all those pain-killers, she does not have reason to believe she will recover, our patient does not thereby reconsider whether she will recover. (I am not sure that, with respect to her own beliefs, she can reason only as a juror, and not also as a detective.) In any case, because I am also inclined toward the stronger thesis mentioned in n. 4, I am inclined to think that whenever a person revises her belief that *p* the person will have exercised evaluative control over her belief that *p* (because, in revising her belief that *p*, the person must have revised her commitments about whether *p*, and, if the stronger thesis is true, the revision of such commitments is accomplished by an exercise of evaluative control). If so, then she will have reconsidered the question of whether *p*. But since I am not prepared to advance the stronger thesis, I will leave it that the agency exercised in revising the belief that *p*, in the case imagined, might remain a bit of a mystery. I will be relatively happy if I have provided a clearer account of at least one way in which reflective control is exercised.

then commands or acts upon the attitude that one has judged unjustified. The unjustified attitude appears, in these metaphors, as an ordinary object of manipulative control. But, on the proposed account, the revision of the belief is not subsequent to the making of the judgment; it is accomplished in arriving at that judgment. Once the judgment has been formed, there is nothing left to command nor anything with which to tinker.

Likewise, if the alternative picture is correct, the powerful picture according to which one's lower-order attitudes conform or are sensitive to one's higher-order judgments (insofar as one is rational) is misleading in the same way. First-order attitudes could not properly be thought of as *sensitive to* the higher-order judgment, because, insofar as one remains of one mind, the first-order attitude will be revised or suspended in the process of arriving at the higher-order judgment. When things go well, the attitude and the judgment do not cohabit the mind. At best, the attitude is sensitive to a stretch of the reasoning that supports or generates the higher-order judgment.

6. CONCLUSION

Given the scope of the topic, my aims have been modest. I hope to have introduced a way of thinking about our agency over certain of our attitudes that I have found fruitful. This way of thinking requires a certain assumption: the assumption that certain of our attitudes embody our answer to a question or set of questions. Given this assumption, it seems we will exercise agency over these attitudes in two distinct ways: by changing our answer to the question(s) they embody or by acting upon them so as to affect them according to our purposes, in roughly the way we can act upon any object that interacts in more-or-less predictable ways with its environment. The first I call exercising evaluative control over the attitude; the second I call exercising managerial or manipulative control.

These two forms of agency are rarely distinguished, because evaluative control does not display the most familiar features of agency while managerial or manipulative control seems to involve an exercise of evaluative control (perhaps more than one). I hope I have suggested why evaluative control deserves to be thought of as a form of agency, despite the fact that it does not sport the usual features. I have also tried to make clear how exercises of managerial control can involve an exercise of evaluative control. Finally, I hope I have shown how certain complex exercises of agency over our minds, including what I have called reflective control, might be modeled in terms of these somewhat simpler forms of agency.[36]

[36] This paper has benefited from the helpful comments and questions of many, including Michael Bratman, Denis Bühler, Tyler Burge, Stephen Darwall, Sean Kelsey, Niko Kolodny, Yannig Luthra,

REFERENCES

Albritton, Rogers (2003) 'Freedom of the Will and Freedom of Action', in Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press), 408–23.

Anscombe, G. E. M. (1957) *Intention* (Oxford: Blackwell Publishing Co.).

Boyle, Matthew, 'Making up Your Mind' (manuscript).

Bratman, Michael E. (1987) *Intention, Plans, and Practical Reason* (Cambridge: Cambridge University Press).

Feldman, Richard (2000) 'The Ethics of Belief', *Philosophy and Phenomenological Research,* 60: 667–96.

Frankfurt, Harry (1971) 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy,* 68/1: 5–20.

——(1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).

Frankfurt, Harry G. (1999) *Necessity, Volition and Love* (Cambridge: Cambridge University Press).

Hieronymi, Pamela (2006) 'Controlling Attitudes', *Pacific Philosophical Quarterly*, 87/1: 45–74.

——(2008) 'Responsibility for Believing', *Synthese,* 161/3: 357–73.

Kavka, Gregory (1983) 'The Toxin Puzzle', *Analysis,* 43: 33–6.

Moran, Richard (2001) *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press).

Owens, David (2000) *Reason without Freedom: The Problem of Epistemic Normativity* (London: Routledge).

Scanlon, T. M. (1998) *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Shah, Nishi (2003) 'How Truth Governs Belief', *Philosophical Review,* 112: 447–82.

Taylor, Charles (1976) 'Responsibility for Self', in Amélia O. Rorty (ed.), *The Identities of Persons* (Berkeley, CA: University of California Press), 281–99.

Velleman, J. David (1989) *Practical Reflection* (Princeton, NJ: Princeton University Press).

——(2000) *The Possibility of Practical Reason* (Oxford: Oxford University Press).

——(1993) 'What Happens When Someone Acts?', in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca, NY: Cornell University Press), 188–210.

Watson, Gary (1982) 'Free Agency', in Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press), 96–110.

Nishi Shah, Sheldon R. Smith; the Philosophy Departments at the University of Konstanz, the University of British Columbia, the University of Pennsylvania, Oxford, the University of Georgia, and UNC Chapel Hill; the participants of the Epistemic Agency Conference at the Université de Genève and the 2007 SPAWN conference; and an anonymous reviewer at Oxford University Press.

8

Trying and Acting

Brian O'Shaughnessy

As a general rule we conceptualize, designate, and single out actions under the general heading, 'the bringing about of phenomenon *x*'. Think of act-descriptions like 'window breaking', 'arm raising', 'jogging the memory'. Here in the first place we mention a particular event, an event which could be described as an 'act neutral' event, in that it can occur either through action or not: for example, in the above situations the events in question are window fracture, arm rise, and an event of recall. Then in the second place we single out a particular act as the active generating of that event. Accordingly, in all such situations, and assuming that *A* is the agent of the action and *x* the act-neutral event, we could say '*A* did *x*'. For example, we see the window break and we say—pointing to that event—'Johnny did that'. Now it is facts of this kind that lead me into a discussion of the expression '*A* did x' in its several uses. I do so because I think we discover an interesting truth when we come to consider mental actions.

1. DISTINGUISHING TWO USES OF '*A* DID *X*'

(*a*) Two Uses of the Expression

(1) We should note that '*A* did *x*' has a simple and familiar use in situations where we are not concerned with actions of any kind. One could call this use the 'causal use', and I shall sometimes refer to it as Use 1. As an example, we point to some flattened corn and say: 'Look what the wind did to the corn'. More exactly, we might indicate some event as causal agency and another event as its effect, and say of the first event: 'Look what that event is doing', for example, 'Look what those hammer blows are doing to that sheet of lead'. And I think we may in general say that when we indicate some entity (wind, hammer, etc.) as agency of a change and some particular state as effect (spread-eagled corn, flattened metal, etc.), we are by implication speaking of such causally related events. This seems to be the case even when we are speaking of causality between states: for example,

between the prevailing unnatural high season temperatures and the untypically fully liquefied state of some lake—say in an Arctic setting. By implication we affirm a causal relation between actual or hypothetical events—whatever else we imply.

(2) I began this discussion by noting the existence of a familiar use of '*A* did *x*': the 'act use', which I shall sometimes refer to as Use 2. This use is of central importance, and I should like at this point to catalogue a few of its main characteristics. One is, that when in this use we say '*A* did *x*', we strictly imply that *A* is a being endowed with the potential for consciousness, and not merely an entity or object of some kind. In addition, in saying '*A* did *x*' we are not automatically to be understood to be citing some specific event of given type, whether in an agent subject or not, that we are assuming to be the cause of *x*, even though we naturally enough believe that *x* has a cause of some kind. Meanwhile, it is illuminating to take note of the range of phenomena to which the act use has application. First, and unsurprisingly, it applies to physical instrumental actions, as when we say 'Johnny did that' of an event of window fracture. Second, it applies to mental instrumental actions, where the instrumentally effected event may be either purely physical or mental in type. For example, on the one hand quickening one's pulse by actively entertaining exciting thoughts, on the other hand creating a state of excitement in one's own mind by the same mental means. In sum, the act use applies unproblematically to instrumental actions of all kinds.

(3) These examples may lead one to believe that in the case of actions we are concerned with a sub-variety of the causal use of '*A* did *x*': namely, that in which the term *A* ultimately refers to an action. However, further consideration shows this to be an error, for it is clear that in Use 2 '*A* did *x*' has unproblematic application well beyond instrumental deeds. For example, it applies to both 'basic' actions and to what I shall dub 'constitutive' actions (where *x* in the first case refers to an immediately effected event in a limb over which one has and has exercised immediate control, and in the second case refers to an event in the natural causal sequence in the body leading up to such an event). Thus, we say in Use 2 'I did that' of an event of arm rise that one has immediately effected, and we say (or should say) in Use 2 the same of an event of biceps-contraction that led up to arm rise in the course of 'basic' action (a claim I shall not defend here).

These facts concerning the range of application of Use 2 demonstrate that Use 2 is not reserved for a sub-variety of Use 1. They demonstrate that the proposition affirmed in Use 2 cannot be analysed into the conjunction of an example of Use 1 and a further claim in which we delimit the type of cause. They show that Use 2 does not affirm a *causal proposition*: rather, it makes an *act claim.* And there is no analytical rule according to which an act claim can be automatically broken down into a causal claim of qualified kind, seeing that the precise nature

of the necessary and sufficient conditions of act-instantiation is a philosophical and not a linguistic issue. For example, it is clear that the truth-value of (say) traditional volitionism, whereby the occurrence of (say) an act of arm raising is understood to necessitate the occurrence of a distinct internal event cause of arm rise, lies within the domain of philosophy rather than linguistics. Only if it were an analytical linguistic truth, comparable to (say) 'all brothers are male', could we subsume the act Use 2 of '*A* did *x*' under the wider head of the causal Use 1: that is, as affirming the occurrence of a sub-variety of the merely causal Use 1 of that expression.

(4) In sum, Use 1 of '*A* did *x*' affirms that an event—or perhaps circuitously an event involving an entity *A*—caused an event *x*. Meanwhile Use 2 of '*A* did *x*' affirms that *A* was the agent of an act that we are entitled to designate as 'the active generating of event *x*': a use which is not to be understood to say that the act caused *x*. Rather, it states that the act is of the type 'active generation of *x*', a designation which leaves it entirely open whether the act in question is 'basic'/'constitutive'/instrumental/mental/physical in kind.

To repeat. When in the act Use 2 we affirm that '*A* did *x*', we are stating that an action occurred which was of the act-type: the active generation of event *x*. We first single out an event *x*—which could even be an act by oneself or some other being, and we then single out an active event which we designate as 'act of bringing about *x*'. Now this implies that the action must be one thing, and the event which the action is the active bringing about of another—the question as to whether or not act and event are distinct phenomena being left open. Expressed slightly differently: the will moves, and an event occurs which in one way or another owes its existence to that movement of the will. It could be an arm movement, a muscle contraction, the appearance in one's mind of a mental image, the fracture of a window, and it could equally be an act on one's own or another's part. But it cannot of course be that very movement of the will.

(*b*) Acting and Trying

This non-identity of the 'movement of the will' and the *x*-event which the action is the active generation of has an important implication. It reveals the existence of a kind of gap which allows for the possibility of *failure*: that is, failure on the part of the will—which in all acts which are 'the active generation of some x' must of necessity be directed towards the generation of some event or other—to lead in whatever way to the occurrence of the latter event. And it simultaneously reveals thereby the possibility of trying and failing, and thus of *trying*. Indeed, in such cases it reveals in addition the *omnipresence of trying*, since nothing could in principle obliterate the sheer possibility of breakdown in whatever link relates the movement of the will and the event to which it is directed. If nothing can as such obliterate this possibility, if in principle the movement of the will may not

lead to the desired event, the movement of the will must be a trying or striving directed towards the occurrence of that event. It is so, however certain one may be concerning the outcome of events.

2. A THIRD USE OF '*A* DID *X*'

(*a*) Explaining the Use

How wide is the application of the above rule? It has universal application in the case of physical actions: that is, actions in which the variety of willing involved is of the bodily kind. It is simply not possible to conceive of an act of the bodily will in which the aforementioned gap is inexistent. Then one might at first suppose that the situation must be entirely different in the case of the mental will, bearing in mind the complexity and heterogeneity of the structures encountered in the case of the mental will. And to some extent this is true. But there can be no doubt that the concept of trying has application in the case of many, and perhaps most, mental or interior 'movements of the will'. Witness the following cases: trying to remember a name, trying to hear a faint sound, trying to call up an image, trying to concentrate, and suchlike. Nonetheless we shall on further investigation discover that the concept of trying has a strangely circumscribed application in the mind. And this conclusion has the important implication that the general rule linking action and trying needs qualification—as we shall discover.

To bring this to light, I will at this point set out a further important use of '*A* did *x*', which I will call Use 3. Like Use 2 this use of the expression is an active use (as opposed to the merely causal Use 1). However it differs from the causal and act uses, both in the values *x* can take, and in the fact that in Use 3 the 'doing' and the 'done' are one and the same. Whereas *x* in the case of Use 2 takes what we called an 'act neutral' value—such as arm rise or window splintering—an event which is non-identical with the act in question—in the case of Use 3 *x* invariably singles out what might be called an active object (a 'doing'), and this object is identical with the aforementioned 'doing'. Examples of Use 3 are the following: 'Churchill did that' where 'that' refers to the act of initiating the invasion of the Dardanelles in 1915. Or: 'The man at the next table did that', where we are referring to 'the fixing of the 1919 World Series' (*The Great Gatsby*).

Now there is one other interesting respect in which Use 3 differs from act Use 2. Whereas in the case of Use 2 we can apply the expression only in situations in which *an action* has occurred, as in 'I did that' where 'that' refers (say) to the movement of one's arm, and are unable to apply '*A* did that' on those occasions when *all that we know* is that a trying to generate some event has taken place, no such restriction is operative in the case of Use 3. Here, in effect,

we are singling out nothing but a 'movement of the will', whether it be a successful or unsuccessful attempt to do some deed, or an act of some kind or another—striven for or not. In this use *x* might stand for an intentional or unintentional action, but might equally apply to a wholly unsuccessful trying to do some deed. In either case a 'doing' of the 'done' took place.

(*b*) The Sense of 'A Did—' in Use 3

What does this tell us concerning the *sense* of claims of the kind of 'A did—' in Use 3? I think it informs us that the expression here affirms that an event of the type of willing occurred, and thus also by implication that an event occurred which was the immediate expression both of an act-desire and an act-intention. Then my claim is that that is *all that it says.* Perhaps the best way to characterize this state of affairs is by saying that 'A did—' in Use 3 states that a 'doing' occurred, where the sense of 'doing' being understood is such that the 'doing' must be active in nature but need not be an action of any kind. Thus, the claim would be in order if all that happened was a completely failed attempt to do some deed. In other words—and unlike the claim being made in Use 2—this present use depends upon the fact that willing can occur in the complete absence of any event (act-neutral or not) of which in Use 2 one could say 'he did that'.

This property of the will obtains because willing need not be the *producing or generating* of anything whatsoever. It is natural to conceive of action—which has such close links with causality—as if it were a sort of 'transcendent causality', a phenomenon in which generation or production occurs, but of a special supra-causal kind reserved for beings endowed with consciousness or mentality. And this is something which finds due recognition in the existence of Use 2 of 'He did—'. But this characterization cannot be extended all the way to willing itself—even though Use 2 and Use 3 apply simultaneously in countless situations, for example—(we might in one and the same situation say 'he did that shattering you can hear' and 'he just did a destructive deed')—and despite the fact that the productive character of willing is for the most part of the essence of the phenomenon, so that a completely failed trying must be accounted an uncharacteristic or even poor specimen of the breed.

3. WILLING AND PRODUCING

(1) Now some might challenge the claim that the will could in principle be exercised when absolutely nothing is produced or generated. They might say that, when we try, the sheer fact that we do try to do some deed amounts to the bringing about of *something,* viz. trying itself. In my opinion, this is confused. Trying *as such* is success of no kind. To be sure, it can be an achievement

to get oneself to even try: 'at least I tried', one sometimes says, and this can be a perfectly natural and valid thing to say. But the achievement was *getting oneself* to try. In and of itself the trying got nowhere, as one might say. One can congratulate oneself for trying, but the congratulations are not over anything the trying was responsible for—not even the trying itself (since, after all, trying bears no sort of responsibility for itself). What is effected or accomplished when it is true that 'at least I tried' is the overcoming of an impediment. This is of the type, achievement: a bringing about—of trying. But to repeat: trying, in and of itself, pure and simple as one might say, gets nowhere. It gets nowhere beyond itself.

(2) In short, the mere fact that when we try we ourselves are responsible for the occurrence of that trying event constitutes no sort of difficulty for the claim that the will can be exercised when nothing is produced: that is, for the claim that its character as a kind of 'transcendent causality' cannot be of the essence of willing, even though that is the natural form taken by the phenomenon. The best example I can think of for bringing this truth home is that of talking to oneself—although other examples exist (as we shall see). Talking to oneself does not consist in the active generating of the imagined sounds of words: a claim I shall not discuss here, since I consider myself to have adequately defended it elsewhere.[1] It will be admitted that the phenomenon in question is of a highly puzzling nature. The temptation to interpret talking to oneself as a form of imagining is very strong, but must I think be resisted. In particular, talking to oneself is not an imagining of the activity of speaking. It is neither an *imagining that* one is speaking (which is a perfectly real but different phenomenon), nor an imagining which takes a direct non-propositional object—an *imagining of* (which equally is a real phenomenon—one which is encountered when (say) in a dream one is making a speech).

In any case, the unusual character of talking to oneself may not be directly relevant to the question we are considering, which is the question as to whether action can occur in the absence of the active production of some event or another. I say so, because there exist phenomena other than internal speech in which action occurs in the absence of production: that is, where it is not open to us to affirm a proposition of the form 'I did *x*' in Use 2.

(3) Now I claimed earlier that whenever a physical action occurs there has invariably to be some event which owes its existence to the bodily will. Another way of expressing the same claim is to say that physical actions can *invariably* be singled out as the active production on the part of the bodily will of some phenomenon *x*, so that mention of the phenomenon *x* must occur in the specification or description of the action. Thus, if *x* is arm rise then the act is an arm raising, if *x* is window fracture the act is window breaking, and so on. If I am right on this matter, and if there exist acts of a non-productive

[1] *Consciousness and the World* (Oxford: Oxford University Press, 2000), 380–2.

kind, those particular acts must be *mental or internal* actions, acts in which the variety of 'movement of the will' is internal—as it is in (say) listening or thinking. Clearly, this cannot be a universal property of the mental will: examples abound of mental actions which are to be individuated as the active production of some phenomenon or other—which may be mental (as in recollecting) or purely physical (as in quickening one's pulse by actively generating imagery). Nonetheless, it seems to me that there exist actions of a non-productive kind, and they prove invariably to be phenomena in which the variety of willing involved is mental in type.

(4) I have instanced talking to oneself as an action in which nothing is produced. Now one might be inclined to attribute this property to the fact that in speaking to oneself one is actively marshalling meanings. But a consideration of the varieties of imagining demonstrates that non-productive action can occur in the absence of the use of a symbolism or language of any kind. For example, a simple imagining-of the performing of a bodily action, where by 'imagining-of performing' I mean something other than 'imagining that it is true that one is performing'. I mean, an imagining in which an experience occurs which is seemingly of the executing of an action (rather than one in which a particular fact seemingly obtains): an experience one would describe as 'seemingly doing physical action *F*'.

Now one might question the reality of such a variety of imagining on the grounds, first that the imagination does not literally reproduce its object, and secondly that willing must be immediately and near infallibly self-identifying. But this must be mistaken. Dreams occur not merely in which a seeming fact is conjured up, but in which an experience occurs whose nature is at the time incorrectly experienced. While dreaming that one is amused may well necessitate genuine amusement, dreaming that one has toothache does not necessitate the occurrence of the experience we undergo when we are aware of a real toothache. And the same is true of the bodily will. In a dream one might seemingly be walking along a street, and no experience of the type bodily willing be taking place. Then even though on such an occasion the imagination is being exercised, we should remember that this imagining is not happening because we will it to do so. It follows that when one dreams one is walking, no willing *of any kind* is normally occurring: neither a mental willing of an imaginative kind, nor any form of bodily willing.

Then it follows in addition that the above phenomenon, in which we dream of bodily willing, cannot be what we have been seeking: namely, an example of action without production. The reason being that action is wholly absent from the scene. The example of non-productive action for which I am searching is rather to be found when the variety of imagining that is occurring takes place in full wakeful consciousness. Thus, I can at will when awake imagine performing a simple bodily deed like moving my arm, and I do so without conjuring up

anything. In particular, when normally I imagine moving my arm, in a special sense of 'seem' I seem to myself to be moving my arm, and this is an immediate and active imagining-of. This situation is to be contrasted with what takes place when I call up a mental image of some object (call it *O*): a procedure in which an element of luck is invariably involved, since as often as not an image instead of a near neighbour or associate of *O* might respond to the call of the will. Now this is definitely not what happens when at will one imagines moving (say) one's arm. And yet—do I not here manage to produce an internal event of willing through imagining willing? I do not. An inflexible rule is, that when one imagines-of an experience *E*, no such *E* is occurring—so that when I dream I am amused and am (normally) really amused in my dream, then I cannot at that moment be imagining amusement in myself. And so it is when I imagine the willing of arm movement.

4. ACTING WITHOUT PRODUCING

(*a*) Imagining

Then if it is true that when I wakefully at will imagine-of (i.e. seemingly experience) moving my arm, no willing is thereby generated in the doing of this willing, *where is the action* of which we speak? Where is the supposed action which I am claiming to be a non-producing? The answer is to be found, not in the object of the imagining, but in the imagining itself. This action, which is of a non-productive kind, is of a wholly mental order: a genuine exercising of the mental will; and in the complete absence of a willing of the bodily variety. Imagining one is moving one's arm is not an occurrence in which one phenomenon (viz. bodily willing/seemingly bodily willing) is either caused by or part of another phenomenon (viz. imagining performing that bodily deed). We do not here—as happens in the case of visualizing—evoke or call up a seeming bodily action. The seeming acting does not *spring forth* in the mind, as imagery does in visualizing: the seeming acting is *one and the same thing* as the mental act of imagining the doing of such a bodily deed. What I experience is my own imagining, not an imaginative effect of that imagining activity. The seeming moving of my arm does not *happen to me,* it does not come forth from my own mind at the behest of my own internal endeavours: it is the endeavour itself. Here there is but one phenomenon that is taking place, and there is but one exercise of the will that is occurring, and they are identical with one another.

Then in the case of such phenomena—and if the feature labelled (by me) 'transcendent causation' is absent—why should we describe them as 'actions'? Why not just call them 'movements of the will'? The reason seems to lie in the fact that they are the complete and full expression of an act-desire and an act-intention: they are willings whose nature is identical with that projected by its progenitors. Thus, I wish or desire to, I choose or decide to, seemingly and

imaginatively in my own mind engage in the moving of a limb, and in the event which ensues I wholly fulfil these several progenitive phenomena. In this sense the 'movement of the will' was successful, and yet nothing in the nature of a 'window' for the occurrence of failure can be detected in such a situation.

(*b*) The Absence of Trying

Then what room is left in this situation for the phenomenon of trying or striving or attempting? None, as it seems to me, and none as we have seen for failing. And yet surely there is a sense in which one might fail here in one's endeavour? One can fail only in a sense that is not relevant to our purposes. Thus, it can on certain occasions be difficult for one to engage in the activity of—to perform an act of—(wakefully) imagining-of (i.e. seemingly experiencing) moving one's right arm. For example, I am in the close vicinity of a road drill, and cannot even (as we say) 'hear myself think'. In this situation the difficulty consists in the absence of the requisite attentive mental space needed for the envisaged imaginative experience: the limited canvas of the attentive mode of (or form of) consciousness is (so to say) completely occupied by a dense paint which cannot be sufficiently diluted to share that space even in the least degree with a prospective brother occupant. Tremendous concentration might on occasion manage to overcome this difficulty, and this must be accounted an achievement, but mostly this is not what happens. It is true that, as a result of the existence of such phenomena as this latter, the concepts of trying and failing must have application in such cases. However, it would not be a *trying to imagine* that is revealed as a possibility. Rather is it a trying to retain the attentive space that allows for the possibility of such an activity.

This trying is of the type, trying to make possible. And we should at this point note the existence of another variety of trying which has application in such situations. I am thinking of what could be called 'trying to get myself to', or (say) 'trying to avoid being deflected away from'. For example, just as I about to imagine moving a limb, a sublime piece of music swims into my hearing, or a juicy piece of gossip coming from the next table does the same. In either case, and just as was the situation in the case of the road drill, there proves to be insufficient attentive space for the imaginative enterprise, although here the cause lies within my own inclinations or tastes rather than from without, where it took the form of intense shock waves impinging upon my auditory apparatus. Here what I must fight is myself! I must not allow myself to be deflected away from the task in hand. And to whatever extent I win that battle, an achievement has occurred: a success as opposed to a failure. A desired effect has been effected, and a correspondingly successful trying has taken place. In such a situation a trying has occurred, but once again it was not a trying to imagine: rather, a trying to *get myself to imagine*, a trying not to be deflected elsewhere, a trying to resist alluring forces—rather like Ulysses and the sirens.

(*c*) The Universal Rule

The property of 'transcendental causation', which is so typical of the will, and universal to the bodily will, is instantiated in these situations, and no doubt in many situations where imagining-of a bodily action is one's aim (or, for that matter, talking to oneself). However, none of these examples of trying are cases of trying to do the deed. They are trying to make the deed a possibility, or trying to get myself to do the deed, and suchlike. But trying to move a limb, or (say) trying to open a door, are not to be confused with a trying to make possible such a deed or trying to get myself to perform it. Thus, I might try to make possible the raising of my right arm by removing with my left arm a mass of debris which has trapped that limb. Equally, I might try to bring myself to ask my irritable and stingy employer for a 'raise', and do so by conjuring up in my mind images of the glamorous holiday a little bit more money would make possible. But these enterprises are to be distinguished from the trying itself.

The claim which I am therefore making is that, in the case of talking to oneself, or (wakefully) imagining-of moving a limb, the will is operative and trying is absent. In other words, the phenomenon of willing is not *as such* a phenomenon of trying, even though the reverse is true. Whereas trying as such is willing, willing is not as such trying. All that willing *as such* is, is *willing*. And when trying does occur, it is both co-present with and identical with a willing. On such occasions they are one and the same phenomenon.

It had seemed at one point in the history of the study of the will that a general rule could be propounded that whenever we act we try or strive to do a deed of some kind or another. And such a rule is valid in the case of the bodily will. But the mental will differs fundamentally from the bodily will on this count. Here the rule breaks down. While there exist many mental willings which are identical with tryings/strivings/attempting to do some deed or another, there are many which are not. The important lesson that the phenomenon of mental willing teaches us through these particular phenomena is that the really universal rule in the case of action must take the following form. Whenever we act the will is operative, we are engaged in a 'doing', and the 'doing' is identical with the action: a psychological event of the kind of 'doing' or 'willing' occurs on all of these occasions; but the willing in question need not take the form of a trying. 'Trying' and 'willing' are not different names for the one phenomenon. This lesson a sub-variety of mental actions teaches.

9

Perceptual Activity and the Will

Thomas Crowther

Reflection on the nature of perceptual activity reveals a type of mental episode with both active and passive constituents. Where one listens to something, watches it, or looks at it, one perceives it. It is a familiar thought that in perception we are passive and at the mercy of our immediate environment. But perceptual goings-on like listening and watching are also active. In this chapter, I explore the nature of this neglected category of the mental through an attempt to understand the dual nature of episodes of this kind. I shall draw on a recent discussion of the role of the rational will in mental activity to offer an account of the structure and nature of a central case of perceptual activity, an account that can explain how active and passive elements are related to one another in this type of episode.

1. AGENCY AND PASSIVITY IN PERCEPTUAL ACTIVITY

Some perceptual goings-on—amongst them listening, looking, observing, scrutinizing, and watching—are active, or 'agential'. I shall set aside for present purposes questions about exactly how the notion of the agential is to be determined and how the boundaries of that category relate to the notion of 'action'. In marking such occurrences out as agential I intend nothing more than that such goings-on are things that agents do, rather than things like digestion or resuscitation that merely go on in agents or that merely happen to them. For one to listen to something, for example, is not merely for something to happen to one or for something to merely go on in one.[1] It is for one to do something.

[1] In this chapter I concentrate discussion on the agential process of listening. Prior to further work, it would be a mistake to simply assume that the account offered of the nature of listening can be transplanted across the sense modalities to do service as an account of watching, looking, or of tactual exploration (feeling). Nevertheless it may be possible to think of the account of listening as a model and to leave it to further work to see how far such a model can be extended. I discuss watching in more detail in Crowther (2009) and offer some further remarks about the relationship between listening and watching below.

One may listen to something carefully, intently, painstakingly, or methodically. That listening is agential does not imply that listening to something is necessarily an instance of 'full-blooded action', or an instance of mental process in which an agent's active powers are exercised to the full.[2] It may be that one listens to something absent-mindedly, or carelessly and without any great attention. Nevertheless, absent-mindedly listening to something, for example, is to be distinguished from something which merely happens to a subject just as much as something done with intention, heed, or care. One can no more digest something absent-mindedly or unintentionally than one can digest it carefully.

Though perhaps not all perceptual activities do so, some perceptual activities entail the perception of the object of that activity. If one listened to the sound of someone talking throughout a period of time, for example, then necessarily one aurally perceived, heard, her. And if one watched a bee flying from plant to plant throughout a period of time, then one saw the bee flying. One could not have watched the bee during that time unless one had seen it.

But hearing something—and, quite generally—perceiving something, is not something agential: hearing something is not something that a perceiving agent does. Different reasons have been offered in support of this claim. Generalizing to a different sense modality the thoughts about seeing offered in the course of the discussion by Zeno Vendler (1957) we could note that one cannot answer the question: 'What are you doing?' with 'I am hearing O' in the way that one can answer that question with 'I am listening to O.' In further support of this claim, Vendler notes that one can't intelligibly qualify perception verbs with the 'agential' adverbial modifiers so readily taken by verbs like 'watching' and 'listening'. One can't intelligibly describe someone as 'hearing something carefully' nor 'seeing something intently'.[3] And we can take these observations about what we can and cannot intelligibly say to be manifestations of the truth about hearing (or seeing) themselves.

These observations place a burden of explanation on the theorist. Given that listening to something cannot be aural perception of (i.e. hearing of) that thing, on pain of its being both agential and non-agential, some explanation needs to be offered of how it can be that perceptual activities like listening to something, watching or listening to something entail the perception of their objects.[4]

[2] My use of the term 'agential' here, and the distinction between full-blooded actions and less than full-blooded actions takes up the three-fold distinction between categories of goings-on suggested in Frankfurt (1988) and Velleman (1989, 1992, 1996).

[3] Vendler (1957) goes on to maintain that seeing (and by extension hearing) are not things that go on or unfold over time at all. They are states or conditions of the perceiving subject. Any such claim, though, I take to be a development of the thought that perception is not agential, and not something one is committed to simply by the observation that it is not possible to hear something carefully.

[4] Brian O'Shaughnessy (1992, 2000: ch. 14), calls this the 'Antitheticality Puzzle'. My discussion here is indebted to O'Shaughnessy's treatment of this puzzle. See ss. 3 and 4 below.

2. AN INSTRUMENTAL CONCEPTION OF ACTIVE ATTENDING

In this section, I consider an attempt to explain this relation in which active attending is to be understood as instrumental in structure. Though this account fails to provide a general explanation of perceptual activity, appreciating the reasons why it does not deepens our understanding of perceptual activity by revealing a distinction between two different categories of perceptual activity. The nature of this distinction is determined by the ways that episodes of these distinct types occupy periods of time. Focusing attention on this distinction will reveal a feature that a good account of the entailment in question will need to possess.

Gilbert Ryle (1949) draws a distinction between 'task' verbs and 'achievement' verbs.[5] Tasks are those things that subjects can be intentionally engaged in, and they are things that necessarily occupy periods of time. As instances of tasks (or activities or processes) he describes: looking for, searching, kicking, running, walking, climbing. In the category of achievements on the other hand are such things as scoring a goal, winning the race, reaching the summit, and finding one's keys, but also seeing and hearing.[6] Though they are occurrences, or events, achievements do not take up time in the way that tasks do: they are instantaneous or durationless happenings that consist in mere changes in, or of, something. In the case of scoring a goal or winning the race, those changes consist just of the termination of a task.

One might be tempted to take the bodily task and achievement pair 'search' and 'find' as a model for understanding the relation between perceptual task verbs on the one hand and what Ryle calls 'verbs of perceptual success' or 'perceptual detection verbs' on the other.[7] In searching for something, one engages in a task or activity with an end or aim that provides its instrumental goal. With such instrumental goals, necessarily, the achievement of that goal, where the achievement of that goal is brought about in the right way by the performance of the task, provides a logical terminus for the task. Looking for one's keys is an activity governed by the goal of finding one's keys. The achievement of the end of that activity, finding one's keys, is an instantaneous change, an occurrence that brings looking for one's keys to a close. Correspondingly, the thought might go, a task like listening to something is an aural task which has hearing or aurally noticing (perhaps hearing a sound) as its end. The achievement

[5] See Ryle (1949: ch. 5). For further related discussion see Ryle (1954: 93–109).

[6] In support of this claim about seeing, Ryle cites, approvingly, Aristotle, who he takes to have said that that 'I can say I see it as soon as I can say I have seen it'. For discussion of this claim see Ackrill (1965).

[7] See Ryle (1949: ch. 7, p. 225; also 1954: 102–5).

of the end of that task, hearing a sound, necessarily brings such aural searches to an end. Successful aural searches are those perceptual tasks, tasks of listening for something, in which the instrumental end of the activity has been achieved. This modelling of active attendings on successful instrumental actions, when conjoined with Ryle's ontology of tasks and achievements, suggests an account of the entailment in question.

The suggestion is that actively attending to something entails perception, because actively attending to something is the successful performance of a task of actively attending, and perception is the instantaneous change in state that constitutes the achievement of the end of that activity.

Though there is much of interest in these suggestions they cannot provide a general model for perceptual activity.[8] Listening to O, for example, is not a process that is, necessarily, terminated by hearing O. One generally hears O throughout a period of time during which one listens to it. Similarly, if one watches or looks at O one generally sees O throughout the time that one watches or looks at it.

The minimal point is not that perceptual searches are not agential perceptual activities. However precisely the details of such an account ought to be developed there are clearly instances of perceptual activity that have (at least something like) instrumental structure; consequently, there are clearly occurrences that we can understand in terms of the success of such goings-on. Rather, we ought to distinguish, prima facie, between instrumental perceptual goings-on—looking for the book, watching out for the eagle, listening out for the bell—and the events which are successful episodes of such activity, and perceptual activities which lack such instrumental perceptual goals, watching the eagle, looking at the book, listening to the bell.

The ground of the distinction between these two categories of perceptual activity is the different ways that episodes of the relevant kinds occupy time. Within the framework of the Aristotelian verb classification offered by Zeno Vendler (1957), Susan Rothstein (2004) distinguishes between 'telic' and 'atelic' goings-on. Occurrences are telic when they have 'a natural stopping point' or when they are 'movements towards an endpoint where the properties of the endpoint are determined by the description of the event'.[9] Goings-on are atelic, by contrast, when 'once they have started, they can go on indefinitely,

[8] There is also much to take issue with independently of concerns over whether the relation between an agential occurrence and perception in cases of perceptual activity can be understood instrumentally. In particular, the denial in Ryle (1949) that seeing and hearing are not processes or things with temporal duration at all (cf. here Vendler 1957) but mere durationless startings or stoppings arguably conflates the fact that hearing is not something that a subject can do deliberately with the fact that hearing is not something that necessarily goes on through time or exists by unfolding over periods of time at all. The conflation here appears between agential process and process simpliciter.

[9] Rothstein (2004: 7).

since the nature of the eventuality itself does not determine its endpoint'.[10] If John ran, then his running is not an occurrence which progresses towards a terminal point, as his building a house does. John's running is an 'activity' or a 'process', while his building a house is an 'accomplishment'.[11] A mark of this difference between kinds of occurring or occurrence is that atelic goings-on are generally 'homogeneous' and telic occurrences are not. According to Rothstein (2004) (assuming here that homogeneous goings-on are the semantic values of homogeneous predicates), 'if a predicate is homogeneous then x P-ed for y time ENTAILS that at any time during y, x P-ed was true.'[12] So, for example, the activity of pushing a cart is homogeneous, in that if John pushed a cart for *y* time then for any time during *y*, John (had) pushed a cart at that time. But it is not the case that if John wrote a letter for *y* time that entails that at any time during *y*, John (had) written a letter at that time. A second mark of the distinction between atelicity and telicity is that: 'If P is an activity predicate (i.e. a predicate that has a semantic value that is an atelic going on), then x is (now) P-ing ENTAILS x has P-ed.' Telic occurrences yield the imperfective paradox in the progressive, but processes or activities, atelic occurrences, do not. If John is now pushing a cart then John has pushed a cart, whereas it is not the case that if John is now writing a letter he has written a letter.[13]

The distinction between the two kinds of perceptual activities distinguished above is a distinction between telic and atelic agential perceptual goings-on. On the account of act-structure just considered, actively attending to O is understood as a telic 'accomplishment' that is not homogeneous. That S actively attended to something, in this sense, during t1–tn, does not entail that S (had) actively attended to O at any time t during t1–tn. If S turned his visual attention on the person behind him from t1–tn, for example, there is a time t, during t1–tn, for example, at which S is only halfway through turning his perceptual attention on the person behind him. But, there is a sense of 'actively attends to' that is atelic. Listening to O (like watching O or looking at O) does not involve a 'progression towards a set terminus determined by the nature of the eventuality itself'. Such goings-on are homogeneous: if S listened to O throughout t1–tn then S (had) listened to O at any time throughout t1–tn.

[10] Ibid. [11] Vendler (1957); Kenny (1963).

[12] Rothstein (2004: 14). See Vendler (1957); Kenny (1963); Dowty (1979); Taylor (1977).

[13] A number of writers dispute that homogeneity and failure to instantiate the entailments distinctive of the imperfective paradox can be taken to be constitutive or explicative of the distinction between atelic processes and activities and telic accomplishments. Taylor (1977) and Dowty (1979) maintain that not all activities or processes are homogeneous. Some processes are only homogeneous down to minimal parts. The thought is that though walking is obviously an atelic activity, it is not the case that if S walked for time t, S walked is true at every time or period of time during t. There may be units of time during which S did not complete the cycle of activities necessary to have walked. I leave these complications aside here.

3. O'SHAUGHNESSY ON THE STRUCTURE OF LISTENING

In an absorbing discussion of listening, Brian O'Shaughnessy (2000) identifies and engages with the issue that concerns us here, and makes some interesting suggestions about how listening can have both agential and passive components.

According to O'Shaughnessy (2000), perception cannot be agential.[14] First, were perception agential, then it would be possible for there to be such a thing as rational hearing, and it would also be possible for us to say that we were engaged in hearing, or the process of hearing. But there is not such a thing as rational hearing, and it is unintelligible to suppose that hearing is something that one could be engaged in doing. Second, like belief, desire, and emotion, the 'will status' of perception (i.e. whether it is agential or non-agential) is fixed by its type. If there are any examples of hearing that are non-agential, then every instance of the kind must be so. But there are undoubtedly non-agential hearings. For one can hear the door slamming non-agentially, or without that being something agentially done. So hearing must be non-agential. Third, hearing cannot be agential if it is to be a source of knowledge about the world.[15] Were hearing agential then the activity of the episode would be explained in terms of causation by act-intentions and desires. But if hearing is to be a way of finding out about the world, the hearing itself cannot be caused by our states of mind, but must be caused by the objects heard. He writes: '(S)urely perception, of its nature and therefore universally, is a responding-to or suffering-of at the hands of its object, howsoever much this may be intentionally engineered by the being who suffers such self-engineered experience'.[16]

But, according to O'Shaughnessy, the relation between listening and hearing cannot be understood as the kind of relations obtaining between the elements of a piece of instrumental action (i.e. the kinds of relations that obtain between perceptual activity and perception in the Rylean model discussed above). While there are elements of the structure of the activity of listening which have analogues in the case of bodily action; in both listening and raising one's arm there is a 'willed event' (hearing in one case and arm rising in the other) and an 'event of willing' ('trying', 'striving', or 'willing to listen' and 'trying to move one's arm') which in cases of successful action are, he claims, constituents of the same event, listening does not involve the activation of a mechanism by which, in successful action, one event brings about the other. In the case of listening,

[14] O'Shaughnessy uses the term 'active' to something that is an expression of agency. I continue to use 'agential' for the sake of continuity.

[15] See Williams (1973); Wittgenstein (1980: 79, 91–4). See Baldwin (2003) for discussion.

[16] O'Shaughnessy (2000: 389–90).

unlike the case of bodily action, there is 'no gap to be mechanistically bridged'.[17] According to O'Shaughnessy, one can try to raise one's arm and yet fail to do so because of the failure of an otherwise extremely reliable mechanism that relates the willing of bodily movement and the willed bodily movement. But by contrast one cannot make sense of trying and yet failing to listen to something. In discussion, O'Shaughnessy turns to the example of attending to a bodily sensation:

> We shall suppose that the project is directing one's attention onto a medium-sized sensation of contact, sited on the heel of one's left foot, for the span of a few seconds. And so one turn's one's attention onto this psychological object, and of course succeeds, and instantaneously. Might a mental mechanism connect these two events? Might a mechanism link the event of turning the attention onto or towards, and the event of noticing or making attentive contact with this object?[18]

For it to be intelligible that there is such a mechanistic relationship, he maintains, it must be possible, however unlikely in fact, for failure to occur, where failure is an unsuccessful attempt to do something and not a simple omission to do. But there is, O'Shaughnessy argues, no possibility of the relevant kind of failure in the case of the kind of mental action in question. One may fail to attend to a bodily sensation through 'changes or abandonment of project or boredom, or sheer forgetfulness. Each of these causal influences might lead, not to the unsuccessfulness of the project, but its termination.'[19] But that is not failure as an unsuccessful attempt to do something. However, can't one try to listen to something, and yet fail, if, say, there is too much interfering noise? 'It is true that we sometimes mean by "trying to listen" trying to keep one's mind on the task of trying to listen, trying for example, to keep one's mind off the deflecting power of alternative attractions, and here of course one can fail—but they are not in question here.'[20] Failure in such cases ought to be understood as a failure of trying to try to listen, not a failure of one's trying to listen.

O'Shaughnessy proposes that listening is to be identified with the agential phenomenon of 'striving-to-listen' or 'willing-to-listen'. Though listening is agential, listening has hearing as a constituent. The hearing (more correctly, given that hearing may go on after listening has stopped, that part of hearing) that is part of listening is to be identified with that part of hearing that is causally explained by the striving to listen. If hearing is to be a way of knowing about the world, then hearing itself cannot be caused by the act-intentions or act-desires of the agent. Listening, or striving to listen, is to be understood as an agential occurrence which is directed at a sound object, which has the causal power to cause it to be the case that such and such a sound object causes hearing of that sound. In listening one sets oneself up to be causally influenced by particular sounds, and the structure of listening is completed by hearing these sounds.

[17] Ibid. 385. [18] Ibid. 385–6. [19] Ibid. 386. [20] Ibid. 386.

The causal power of a will-to-listen is in the nature of an attractive power, and its presence is determined by choice. Freely selecting whichever feature interests us, say the timbre of the sound, we overtly open the door to timbre's causal influence upon the attention. And we actively do so. We actively make the attention open to the influence at the hands of timbre. We do what deflects any occurrent hearing in this direction, thereby ensuring that the attention tends to light upon timbre. In this regard, the act is not unlike an instrumental act—though strangely in reverse. Thus, we actually enlist the timbre of the sound as an external cause of timbre-hearing, through specifically 'grooming' any possible hearing for timbre-affectedness. This trying, that the attention be thus influenced, is the continuous inner creation of a causally influential internal mould; and as the desired effect of this continuous mental work occurs, which is to say a hearing of the desired kind, an act-edifice is completed, namely listening to the heard feature of the sound.[21]

According to O'Shaughnessy, listening is agential, because listening is striving to listen, or willing to listen, and striving to listen or willing to listen is something that one does. Hearing is not agential. It is not possible to hear rationally, there are some non-agential hearings, and hearing plays an epistemic role. Listening entails hearing because to listen to something is for one to strive to listen to a sound and for one's striving to listen to a sound to cause one's being caused to hear sounds by the presence of those sounds.

4. DIFFICULTIES FOR THIS ACCOUNT

There is a good deal in the discussion of O'Shaughnessy (2000) that we should see our way to preserving and further explaining. But what I want to concentrate on here are some worries with that account that emerge when we consider whether what is presented there is a unitary conception of the act-structure of listening.

In the first instance, the notion of agential perceptual events manifested in the discussion of the direction of attention onto bodily sensations quoted above appears to be instrumental, or at least capable of being glossed in terms of mechanical explanation. O'Shaughnessy argues that there is no way to make sense of the idea of an unsuccessful attempt to direct one's attention onto a bodily sensation. There cannot be a mechanistic connection, he writes, between the 'event of turning one's attention towards a bodily sensation' and the 'event of noticing or making attentive contact with it'. But that is by no means obvious. It may be that one turns one's attention towards a bodily sensation, perhaps in response to a request to concentrate on the feeling of one's foot resting on the floor, but that before one notices the feeling or makes attentive contact with it, the sensation is not available to be attended to as a result of the action of some fast acting anaesthetic the administration of which is caused by changes in

[21] Ibid. 397–8.

the locus or disposition of one's bodily attention. Here it seems that failure to make attentive contact with a bodily sensation is not simply an abandonment of the perceptual project or an omission. If the example is indeed a model for the relationship between an event of willing and an event that is willed in the case of an event of listening to a sound, then failure to listen is conceivable for related reasons. One may be instructed to listen to the sound that the neighbours are making, but before one notices or makes attentive contact with that sound, the sound is no longer available to be perceived. There is a clear sense here in which we can make sense of the possibility that a subject engages in willing to listen to a sound but nevertheless fails to make aural contact with any such sound. This conception of listening, though, is subject to the worries about conceiving of the act-structure of listening as instrumental or telic set out in the previous section.[22]

Elsewhere, though, O'Shaughnessy suggests a different picture of the relation between perceptual activity and perception.

> The situation is like this. A sound exists, or perhaps a sensation, of which we are already aware, and for some reason we choose to actively attend to it. And so a will-to-listen takes place, that proves to be a successful listening, in relation to which the sound is at once immediate material object and immediate unrationalized part-cause of the hearing-sector of the same listening. (2000: 392)

Suppose instead we take the conception of the relations between the event of willing and the event which is willed manifested here as our model for understanding the process of listening and its relation to hearing. Given this conception, 'awareness of the sound' or awareness of a sensation is a condition of listening taking place at all (or beginning to take place) and taking 'awareness of a sound' to be (or at least to entail) the occurrence of an event of hearing a sound, here it is a necessary condition for the agential phenomenon, striving to listen, to take place, that the event of hearing already goes on.[23] The image

[22] Another aspect of the account presented in the discussion of active attending in O'Shaughnessy (2000) that encourages a view of listening as instrumental are the schematic diagrams (pp. 390, 392) of the preferred theory of the act-structure of listening. The diagrams suggest that, at least for some events of listening, there is an initial period of time during which the agent strives to listen but does not hear. If that is a possibility, then singling out a particular event as an event of the type striving to listen does not entail that it is an event of the type hearing. But then it remains to be seen how, within such views about the agential process of striving to listen, one could avoid thinking of listening as instrumental; as a successful attempt to hear or to produce hearing. (It is consistent with this that singling out a particular event that is a successful striving to listen, i.e. a listening, as an instance of the type successful striving to listen, or as an instance of the type listening, entails that such an event is also, necessarily, a hearing and not a hearing in virtue of an instrumental or merely productive relationship between that event and a hearing event.)

[23] It might be objected to the tack that I have adopted above on O'Shaughnessy's quote about attending to bodily sensation that some such view must be supposed there also. For it is incoherent to suppose that one could have a bodily sensation without being aware that one has it or without already being in 'attentive contact with it'. I agree that there may be no such a thing as directing the attention onto a bodily sensation without the agent already being aware of that sensation. But this point shows nothing more than that the case of attention to bodily sensation may not be the best

apt for perceptual activities like listening here is that of the concentration or focusing of the beam of the attentive spotlight to a greater degree of intensity onto something that is already an object of awareness.

Taking the perception of a sound to be a necessary condition for listening to a sound, or for an event of listening to a sound to begin, certainly makes it more difficult to make sense of failed attempts to listen, where a failed attempt to listen is the occurrence of an event of striving to listen in the absence of the willed event of hearing. But if the relation between agential perceptual activity and perception is understood in this way it is now more difficult to keep clearly in sight the sense in which, as O'Shaughnessy says, the 'aim of listening' is 'the production of hearing', or that the function of listening is 'the generation of hearing'.[24] How could the function of listening be the production or generation of hearing if being in a position in which the event the occurrence of which is the object of such a function (the event the occurrence of which is the object of the 'act-desires' or 'intentions' with which listening is engaged) is a condition for the occurrence or onset of the agential process itself?

These features of O'Shaughnessy's discussion help us to focus the following question: if an activity like listening is not to be understood in terms of a successful attempt to notice a sound or to make attentive contact with (i.e. come to hear) a sound, and if hearing a sound is a condition for the onset of listening, what remains of the intuition that the role of listening in the mental economy is to produce hearing?

5. LISTENING AND THE STRUCTURE OF ACTIVE ATTENDING

In this section I present an account of listening and show how it helps us to explain the relation between agential and passive components of perceptual activity. Though related to a position suggested by discussion in O'Shaughnessy (2000) in the course of his treatment of listening, the account offered here differs in certain crucial respects.

In his account of listening, O'Shaughnessy (2000) takes the basic active phenomenon to be investigated to be 'listening to a sound S'. But one could

way to illuminate a perceptual activity such as listening, where such a distinction (between directing the attention onto a sound, and noticing a sound or making attentive contact with a sound) can be made intelligible. In discussion above I simply take at face value that the example is intended to provide a concrete illustration of a type of act-structure that is present in an activity like listening, an act-structure that includes as constituents an event of willing (an event of 'directing the attention onto' *x*) and a willed event (an event of 'noticing' or 'making attentive contact with' *x*). Of course, there are other, far more basic reasons, centring on the difficulty in accounting for bodily sensation in terms of perception of sensation, for being cautious about using an example involving bodily sensation to shed light on specifically *perceptual* activity.

[24] O'Shaughnessy (2000: 395).

not listen to a sound S without listening to the producer or the source of the sound; without listening to what makes that sound. One could not, for example, listen to the sound of talking without listening to the producer of that sound. The producer of a sound may be a familiar material object, a collection of such objects, or it may not be a material object at all. If one listens to the sound of thunder, one listens to a storm-cloud producing thunder. In general, though, one could not listen to sound S unless one were listening to the producer of the sound S, call it O, making that sound. Listening to a producer of sound is basic in this sense with respect to 'listening to a sound S'.

To listen to the producer of a sound, O, is a process of agentially maintaining aural perceptual contact with (i.e. hearing of) O with the aim of knowing what sound O is producing. Let me treat the different elements of this account in turn.

Listening to a sound-producer O is a process as opposed to an accomplishment or a time-occupying perfective event; it is an atelic time-occupier. First, listening to O is homogeneous. If one listens to O from t1–tn, then one listened to O at any time throughout t1–tn. Second, listening to O exhibits the pattern of entailments that distinguishes it from accomplishments or time-occupying perfective occurrences with respect to the imperfectivity paradox. If S stopped listening to O at t, then S had listened to O.

For one to listen to O throughout a period of time involves one maintaining aural perceptual contact with (hearing of) O throughout that time. For a process to involve the maintenance of aural contact with a producer of sound is a kind of occurrence or going on that plays a preservative role with respect to the perceptual state or condition of hearing the producer of sound throughout a period of time. It is a process of maintaining or sustaining the hearing of O through that time. For listening to be a process of maintaining or sustaining hearing is not analysable in instrumental terms. Listening to O is not a process that causes or brings about the preservation of a state of aural perceptual contact with sound-producer O throughout a period of time. For one to listen to O throughout some relevant period of time is for one to be doing something such that it is in virtue of doing that that one maintains such a state of aural contact with O. In the same way that the steady emission of energy of a certain object throughout a period of time does not cause or bring about the fact that the object is in a condition of radioactivity throughout that time so, I propose, does listening to an object involve the maintenance of a condition of aural contact with O. I shall return to some of the complications that surround this claim presently.

Listening to O is an agential process of maintaining aural perceptual awareness of O. Not every aural occurrence that involves the maintenance of the state of hearing O throughout a certain period of time is agential, is something that the subject does, or can be said to do. If I heard the fireworks exploding for a time then a process went on in me that preserved in me the state of hearing the fireworks, or hearing the fireworks exploding. Nevertheless, if I heard the

fireworks exploding for a time my so hearing them is not something I could have done with care, deliberation, or absent-mindedly (though if I listened to them that is certainly something that I could be capable of doing carefully, deliberately, or absent-mindedly).

Where a perceiving agent, S, listens to a producer of a sound, the aim of that process of maintaining awareness of the producer of the sound O is to provide that agent with knowledge of what sound the producer of the sound is making; to put S in a position in which he knows what sound O is making. For S to know what sound O is making where that state realizes the aim of listening is for S to know that *p*, where '*p*' is the answer to the question: 'What sound is O making?'. The aim of listening is realized where S aurally perceptually monitors O with the aim of knowing what O is doing, and S is in a state of knowing what sound O is making in virtue of the process of aurally monitoring the sound producer O that he is engaged in. It may be, for example, that the aim of listening to O is realized in the knowledge that O is singing, or that O is not making a sound at all. Where the aim of listening is realized in S's knowledge, at t, that *p* (where '*p*' is an answer to the question: 'What is O doing?') and where this state of knowledge is a state that the subject is in, at t, in virtue of the fact that at t he is engaged in aurally monitoring O, we can characterize that condition as one in which S can *hear what* sound O is making, where to hear what O is doing is to *hear that p*, where 'p' is an answer to the relevant question about what O is doing.[25] The aim of listening to O, then, is to hear that *p*, in this sense.

S may listen to O throughout a period of time even though the aim of that process is not realized, or not realized to any interesting or significant degree. If S listens to what sound O is making, it is not necessarily the case that S is in a position to hear what O is doing, at least where that knowledge is anything more than that O is producing *that* sound or that O is making *some kind* of sound. S may have no grasp of the kind of sound that O is producing. S may not know, for example, that O is producing a C, because he does not know what a C sounds like (at least, not where knowing what a C sounds like is understood as knowing that a C sounds like *that*). Alternatively, though S is listening to O, it may be that contingent constraints produced by the circumstances prevent S from knowing, for the majority of sounds that O is producing, that O is making that sound (in virtue of maintaining aural awareness of the producer of that sound). If the agent is genuinely listening to the producer of the sound, though, the process must put him in a position to know what sound O is making for at least some sound that the producer of that sound makes. One who listens to someone who can be seen, through the walls of a transparent sound-proofed

[25] I do not assume here that if S listens to what O is doing, S himself is necessarily in a position to answer the question: 'What is O doing?'. S may be a creature that does not have the ability to speak, and, hence, which does not have the capacity to raise and settle questions about what O is doing.

room, may not be able to hear that he is screaming, but is able to listen to him, for example, because he can hear his feet tapping on a stretch of floor that is not sound-proofed.

Where a perceiver listens to a producer of a sound, what puts that agent in a position to know what sound the producer of the sound is making is the agent's hearing the sound-producer Φ-ing (where 'Φing' can be an answer to the question: 'What sound is O making?') Listening to O is the kind of process the aim of which may be realized in knowledge of what sound O is making because for as long as an agent listens to O, that agent can hear O Φ-ing.[26] We have noted the thought that the knowledge-generating function of certain aspects of our mental lives is inconsistent with the operation of those functions being subject to the will. But there are no such problems in thinking that listening—or active attending in general—is unsuitable for playing a knowledge-generating role (a role that I take it to be essential to its functioning). For while listening is an active agential project, what one is in a position to know via listening (i.e. what one is in a position to hear is the case) is determined by whether one hears O, for example, Φ-ing or Ψ-ing.[27] And that is not subject to the will.

On this account, then, listening to a sound-producer is agential: it is something that the agent does. But hearing O is not agential. Here, we can agree with both Vendler (1957) and O'Shaughnessy (2000) as we distinguish between different senses of 'hearing O'. On the one hand, we might take 'hearing O' to be the name of a state or condition of a subject, a way that the subject can be. Hearing O in this sense is not something that can be done because states or conditions do not have the right temporal shape to be done. One can no more do or engage in hearing that one can engage in red. On the other hand, 'hearing O' might be understood as a perceptual occurrence. In this sense, we take hearing to be a processive constituent of the stream of consciousness. To distinguish it from the stative notion we might refer to such a process as 'aurally apprehending O' or the 'aural apprehension of O'. Aurally apprehending the fireworks exploding is not a state but a processive occurrence that unfolds over time. But it is not something that can be done. One cannot aurally apprehend the fireworks carefully or

[26] The account offers a number of ways to make sense of *listening to O Φ–ing* (e.g. listening to O producing a sound or making a sound). A natural suggestion is that listening to O Φ-ing is what occurs when an agent listens to O, i.e. maintains hearing of O with the aim of knowing what sound O is producing, and where the agent hears O Φ–ing in virtue of maintaining that contact. Alternatively, it may be that an agent listens to O Φ–ing in the sense that it is the Φ–ing itself that is listened to and is the object of that activity. Here, an agent listens to O Φ–ing if it is *O's Φ–ing* that he agentially maintains aural awareness of throughout a period of time.

[27] I think that this is how we should make best sense of O'Shaughnessy's claim that 'hearing completes the act-structure of listening'. Hearing does not complete the act-structure of agential monitoring because it is a conceptual truth that hearing is entailed by the agential maintenance of aural perceptual contact with O throughout some period of time. However, what is true is that merely aurally apprehending O Φ-ing, as one listens, completes the act-structure of listening in the sense that it puts one in the position to realize the aim of the activity of listening, i.e. knowing what O is doing in virtue of hearing O (i.e. hearing what O is doing).

deliberately, or absent-mindedly. Nor is one's aurally apprehending the fireworks something that can be described as something intentional, as something that one tries to do, or something that one wills to do.

Given the account of listening offered here, listening entails hearing because to listen to O is to agentially maintain aural perceptual contact with (i.e. hearing of) O throughout a period of time with the aim of knowing what it is doing. But a subject cannot have engaged in a process of preserving or maintaining perceptual contact with O throughout some period of time without there being some condition of hearing being maintained.[28] In contrast to the account of structure that O'Shaughnessy (2000) offers, on the account suggested here, there's no *productive* relationship between listening and hearing. Listening does not produce the occurrence of an event of hearing or the unfolding of a process of hearing; listening is not the *production* of hearing. Listening entails hearing because listening is the agential preservation or maintenance of hearing.

The relationship between listening to O and hearing O is distinguished from the instrumental conception of their relationship discussed in section 2, though. Listening does not entail hearing because listening to O is an accomplishment, a time-occupying event that has an end or aim which is its telic point, that is, an occurrence that necessarily constitutes the point of termination of such a going-on, and listening involves the successfully bringing about, i.e. causing, of such an occurrence. Generally one hears O throughout the time that one listens to O. Hearing O, therefore, cannot bring a listening to O to an end.[29] By contrast, listening out for O is a telic occurrence that has hearing O as a telic point. Necessarily (barring certain failures of knowledge, self-knowledge, and rationality) if one comes to hear O one cannot continue to listen out for O. Characteristically, listening out for O can be a failure. One can listen out for O throughout a period of time, and yet one's listening is a failure because one failed to hear O, one failed to achieve, through one's efforts, the telic point of that kind of perceptual activity. By contrast, if one listens to O throughout a period

[28] In virtue of the fact that listening to O entails that there is a state of hearing O being preserved, then it also follows that if S listened to O then S heard O in the sense, now, of a perceptual occurrence: that of aurally apprehending O. For if S was in the condition of hearing O throughout a period of time, it follows that S aurally apprehended O throughout that period of time. Hearing O is a condition that the perceiving subject can only be in, in virtue of the fact that a process of aurally apprehending O goes on in him.

Here is not the place to defend this claim in detail. But I agree with a strategy pursued by O'Shaughnessy (2000: 43) here. The relations are manifested in the impossibility of hearing in conditions of a total 'mental freeze' a condition involving the complete absence of any mental activity or processive mental functioning. It is quite possible that one could continue to know certain facts or believe things in such conditions. But in the absence of mental process one could not be in the state of hearing.

[29] It is clear that any such idea would end paradoxically in listening not being able to begin. Given that listening to O is to sustain a condition of perceptual contact with O, starting to sustain such a condition entails one is in it. Were hearing the telic point of the process, a necessary condition for the onset of the process would be a sufficient condition for its termination.

of time, one's listening to O cannot be a failure in the sense that it is a failure to bring about some occurrence that constitutes its telic point.

One's listening to O may be a failure in the sense that one may fail, in one's listening to O, to realize the aim of the process of listening to something. One may listen to O and, for whatever reason, fail to know what O is doing. But again, and for reasons related to those I have just presented, S's failure to come to know what O is doing is to be distinguished from a failure to achieve the telic point of such a process. Generally, at least in those well-placed and in good environmental and mental conditions, one knows what O is doing and knows what O is doing in virtue of the fact that one listens to O, throughout the time that one listens to it. The event which is the acquisition of such knowledge cannot be the necessary point of termination of listening to O, in the way that coming to know that the PM is resigned is the necessary point of termination for one's listening out for information from the report on the radio about whether the PM has resigned. One's failure to realize the aim of the process of listening is also to be distinguished from one's failure to listen. If one agentially preserves aural contact with O with the aim of knowing what O is doing, then one listens to O, whether that aim is realized or not.

Here is not the place to take up the question of how this account of the structure of listening generalizes to other cases of perceptual activity. It is plausible, though, that a good explanation can be given of a central core of the varieties of perceptual activity in terms of the kind of account of listening offered here. Elsewhere, I have argued that we ought to understand watching O as a case of watching what O is doing, where watching what O is doing is visually agentially monitoring O with the aim of knowing what O is doing.[30] Interesting distinctions, I suggest, will be found amongst varieties of perceptual activity, where those varieties of perceptual activity are distinguished by their different epistemic aims. I think it is plausible, for example, that looking at an object or scrutinizing an object is not to look at what that object is doing, but to monitor that object with the aim of knowing (at least standardly) what F that object is (what colour?, what shape?, what kind?). Those, however, are issues to be treated more fully elsewhere.

6. ATELICITY AND THE AIMS OF AGENTIAL PROCESSES OF PERCEPTUAL MONITORING

In this section I want to draw attention to some things that I think stand in the way of a satisfying grasp of the structure of active attending. At the heart of the argument in this chapter has been a distinction between atelic processes and telic

[30] I offer arguments for this in more detail elsewhere. See Crowther (2009).

occurrences. But it is of some note that on very many of those occasions in which one or other of such mental activities goes on, we find instances of the other activity going on, too. The subject who spends the five seconds after the hooter listening out for the sound of the nearby tower-block being detonated then goes on to listen closely to the tower-block collapsing. This, combined with the fact that we are inclined to describe both of the distinct kinds of mental activity going on here—listening to and listening out for—as 'listening' can tempt the assimilation of the different ways that these goings-on have to be explained to one another. But we ought to resist that temptation, for reasons I have argued for.

Similarly, the proximity of these different types of going-on to one another can prevent one from being properly sensitive to the differences between these two activities on the score of their productive role. The denial that there is a causal or productive relation between what perceptual activity one engages in agentially, and the maintenance of perceptual contact with that object throughout a period of time can seem odd. Suppose one is indoors with the curtains closed listening to some fireworks going off intermittently. Suppose also that the fireworks are being let off a long way away, so that they are only faintly audible, and that one has to concentrate hard to distinguish the fireworks from the sound of the television. Imagine that after the first faintly audible bang, one comes to listen to those fireworks for a period of time, and that during the time one listens to the fireworks one hears them bang faintly, say, ten times. Given these circumstances, the thought may go, it cannot be that there isn't a productive relationship between listening to the fireworks and hearing the fireworks. Suppose that after the initial faint bang (i.e. the initial aural apprehending of the fireworks) that attracted one's attention and which constituted the onset of listening, one hears a second bang. The fact that the TV is blaring and that the fireworks only make a faint sound entails that one would not have heard the second firework bang unless one had been listening to the fireworks. But that is just for one's listening to the fireworks to have brought one to hear, or to come to hear the firework banging. This second event of hearing (like the other eight) is brought about or produced by one's listening, by the fact that one has trained one's aural attention in such a way that one was not, say, distracted by the gunshots on the television. So one's hearing of the fireworks is, apparently, produced by one's listening to them.

The objection might continue that this is not only a feature of active attending in the aural modality, but processes of monitoring across other sense-modalities. Think about visually tracking the sparrowhawk that darts through the dense wood one walks through. As one watches the bird dart along, quickly changing its direction and height in flight, one sees it intermittently as it passes behind the birches. But one's glimpses of the bird as it comes out into the open are surely glimpses caused (brought about, produced) by the way that one is training one's visual attention in watching it. It is one's watching of the bird that causes the glimpses of the bird, and given that one's coming to see the bird again is

to maintain contact with the bird, one's maintenance of visual contact with the bird is surely caused by one's watching it.

But this line of argument ought to be resisted. The examples show a number of things. What they show, first, is not that hearing can be brought about by listening to O, but that hearing can be produced by *listening out for O*. Where one came to hear the second bang of the fireworks one came to hear it because one was listening out for the firework banging, one was doing something that involved a training of one's attention in such a way as to make sure that one notices the fireworks if they continue to go off. But that is not to listen to the firework banging. What the example shows, secondly, is that in some kinds of circumstances, for S to listen to O throughout a certain period of time entails that S listened out for it during periods of that time. Let me explain. It was an assumption of the examples that S listened to the fireworks throughout t1–t10. For S to listen to the fireworks throughout that time is for S to agentially maintain perceptual contact with the fireworks throughout that time. That is correct, for the following reason. In the case discussed, S did agentially maintain aural perceptual contact with the fireworks throughout that period of time because though there was a period of time during t1–t10 in which S could not hear the fireworks, he nevertheless heard the fireworks banging throughout that period of time, in virtue of the fact that that period of time is a sub-interval of time during which he heard the fireworks banging. In these circumstances (i.e. circumstances in which the object of the activity goes out of range of the relevant perceptual faculty for periods of time), the fact that an atelic process of agentially maintaining contact with O goes on throughout a period of time entails that during that time a telic event, an accomplishment (successfully listening out for the fireworks) also occurred.[31] It is not one's listening to the object throughout some period of time that brings about the hearing of it, or produces the hearing of it. One's listening to it is to maintain such contact with it throughout the relevant period of time. But that one did listen to an object in such a way can entail that one also throughout that time successfully listened out for it, an accomplishment that apparently needs to be understood in terms of the causation of an occurrence by an instrumental attempt to do something.

What the examples show, thirdly, is that though one cannot listen to some producer of sound throughout a period of time without having heard that thing, it does not follow that one necessarily heard it throughout that period of time. For during the period of time when the fireworks were banging intermittently, there was a period of time during which one could not hear the fireworks (though one throughout that time heard them banging intermittently). In general, for a perceiving agent to have heard O Φ–ing throughout a period of time does not

[31] The same goes for the case of watching, too: one watched the bird throughout the relevant period of time, and given that one did then one successfully watched or looked out for it, during periods of time, too.

entail the agent heard O throughout that time. One may listen to O throughout a period of time, although throughout it one hears O falling silent, or hears O pausing between notes, or hears O not making any sound at all.

Though there are cases in which atelic processes of listening to producers of sounds entails successful aural telic accomplishments, we ought to remain sensitive to the distinction between these two categories.

7. CONCLUSION

The agential and passive aspects of an activity like listening are integrated in the fact that listening to an object is an agential process in which a condition of aural perceptual relatedness to some object is preserved or maintained with the aim of putting the subject in a position in which he knows what sound that object is making. For a subject to listen to an object requires that they hear it, else there is nothing that has been maintained (even if that does not entail that a subject who listened to O throughout a period of time heard it throughout the entirety of that time). This leaves a good deal about the nature of perceptual activity yet to be settled. It remains to be seen how perceptual agency itself is to be explained and understood. But there is no structural problem about how a mental activity can be something perceptual and something agential.[32]

REFERENCES

Ackrill, J. L. (1965) 'Aristotle's Distinction between *Energeia* and *Kinesis*', in R. Bambrough (ed.), *New Essays on Plato and Aristotle* (London: Routledge).

Baldwin, T. (2003) 'Perception and Agency', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness* (Oxford: Clarendon Press), 188–200.

Crowther, T. M. (2009) 'Watching, Sight and the Temporal Shape of Perceptual Activity', *Philosophical Review*, 118(1): 1–27.

Dowty, D. (1979) *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ* (Dordrecht: D. Reidel).

Frankfurt, H. (1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).

Kenny, A. (1963) *Action, Emotion and Will* (London: Routledge).

Mourelatos, A. (1978) 'Events, Processes and States', *Linguistics and Philosophy*, 2: 415–34.

O'Shaughnessy, B. (1963) 'Observation and Will', *Journal of Philosophy*, 60/14: 367–92.

——— (1992) 'The Diversity and Unity of Perception and Action', in T. Crane (ed.), *The Contents of Experience* (Cambridge: Cambridge University Press), 216–66.

[32] I gratefully acknowledge much helpful discussion with Matthew Soteriou and Guy Longworth. Thanks also to an anonymous reader for OUP for helpful comments.

——— (2000) *Consciousness and the World* (Oxford: Clarendon Press).
Rothstein, S. (2004) *Structuring Events: A Study in the Semantics of Lexical Aspect* (Oxford: Blackwell).
Ryle, G. (1949) *The Concept of Mind* (London: Penguin).
——— (1954) *Dilemmas* (Cambridge: Cambridge University Press).
Sibley, F. N. (1955) 'Seeking, Scrutinizing and Seeing', *Mind*, 64/256: 455–78.
Steward, H. (1997) *The Ontology of Mind: Events, Processes and States* (Oxford: Oxford University Press).
Taylor, B. (1977) 'Tense and Continuity', *Linguistics and Philosophy*, 1:199–220.
——— (1985) *Modes of Occurrence: Verbs, Adverbs and Events* (Aristotelian Society Series, 2; Oxford: Blackwell).
Velleman, J. D. (1989) *Practical Reflection* (Princeton, NJ: Princeton University Press).
——— (1992) 'What Happens When Someone Acts?', *Mind,* 101: 461–81.
——— (1996) 'The Possibility of Practical Reason', *Ethics,* 106: 694–726.
Vendler, Z. (1957) 'Verbs and Times', *Philosophical Review,* 67:143–60 (repr. in a revised form in Vendler 1967).
——— (1967) *Linguistics and Philosophy* (Ithaca, NY: Cornell University Press).
Williams, B. (1973) 'Imagination and the Self', in *Problems of the Self* (Cambridge: Cambridge University Press).
Wittgenstein, L. (1980) *Remarks on the Philosophy of Psychology*, ii, ed. G. H. von Wright and G. Nyman (Oxford: Blackwell).

10

Mental Action and Self-Awareness (II): Epistemology

Christopher Peacocke

We often know what we are judging, what we are deciding, what problem we are trying to solve. We know not only the contents of our judgements, decidings, and tryings; we also know that it is judgement, decision, and attempted problem-solving in which we are engaged. How do we know these things?

Such pieces of knowledge are members of the wider category of knowledge of our own mental actions. My aim in this chapter is to give a philosophical account of the nature of our knowledge of our own mental actions. Any account of this knowledge has to dovetail with a theory of the nature of mental action itself. The account of mental action on which I will be drawing for this purpose is one that endorses these principles:

- Mental action is a genuine subspecies of action in general. The differences between mental action and bodily action are fundamentally only the differences between the mental and the bodily.
- You can be aware that you are performing a certain action without perceiving that action, and without bodily perception from the inside of the motions involved in the action. This distinctive action awareness exists for mental action as well as for bodily actions. Perhaps someone will insist that action awareness is simply another form of perception, a form we should recognize in its own right. I would dispute that, but for present purposes we do not need to enter that discussion. What ought to be uncontroversial, and all that matters for the position I will be developing, is that you can be aware that you are doing something without perceiving your action in any of the ordinarily recognized senses of vision, touch, proprioception, hearing, taste, or smell. You can be aware that you are raising your fully anaesthetized arm without any feeling in the arm, and whilst looking in the other direction.[1]

[1] The question of what it is for an event to be a perceptual experience is interesting, potentially significant, and, to the best of my knowledge, underdiscussed. One difference between action awareness and perceptual awareness is that action awareness does not involve sensational properties

- The content of such action awareness is first-personal and present-tensed. It has the form 'I am doing such-and-such now'.

This conception of mental action is one for which I argued in an earlier paper, 'Mental Action and Self-Awareness (I)'.[2]

We need to subdivide the epistemological issues about knowledge of our own mental actions. A first set of questions concerns the nature of the way (if any) in which we come to know what mental actions we are engaged in. What is this way, and what gives it the status of a way of gaining knowledge?

Once we have proposals for answering these initial questions, we have to engage them with some issues that have been central in recent debates about self-knowledge. Many recent treatments of self-knowledge have, rightly in my view, rejected perceptual models of self-knowledge, for a variety of reasons. The question arises: do the proposals I offer fall to the same objections as those to perceptual models of self-knowledge, and if not, why not?

Another fundamental and continuing contemporary issue I will address is that of the consistency of our distinctive self-knowledge with externalism about intentional content. The account of knowledge of our own mental actions I offer makes possible an answer that addresses the concerns of those who think that certain kinds of account of this distinctive self-knowledge cannot be reconciled with the externalist characteristics of intentional content.

These are by no means all the issues that arise in the epistemology of mental action; but any account that cannot address these issues satisfactorily would be a non-starter. I take them in turn.

1. WHAT, IF ANYTHING, IS THE WAY IN WHICH WE COME TO KNOW OF OUR OWN MENTAL ACTIONS?

The distinctive way in which a subject comes to know of his own mental actions is by taking an apparent action awareness at face value. You judge that it will rain. When so judging, you have an apparent action awareness of your judging that it will rain. It seems to you that you are judging that it will rain. By taking this awareness at face value, you come to know that you judge that it will rain. In another case, you may have an apparent action awareness of your

in the way all genuine perceptual experience does. (There is some preliminary discussion in my paper 'Sensational Properties: Theses to Accept and Theses to Reject', *Revue Internationale de Philosophie,* 62 (2008): 7–24, in a special issue on the Philosophy of Mind, ed. J. Proust.) Another difference is that perceptual awareness provides objects of attention, and action awareness does not. It is a question whether these differences are fundamental, or are rather by-products of something else that is more fundamental. As these tentative remarks suggest, the whole issue merits further consideration.

[2] In J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind* (Oxford: Blackwell, 2007).

calculating the sum of two numbers; by taking this awareness at face value, you come to know that you are engaged in calculating the sum of two numbers; and so forth.

Apparent action awareness is a belief-independent event. A thinker may or may not endorse in judgement the content of an apparent action awareness. An amputee may know very well that if he tries to raise his missing right arm, he will have an apparent action awareness of raising it, even though he is not in fact raising it. This subject will not endorse the content of his apparent action awareness. Because action awareness is not the same as judgement or belief, a self-ascription of an action made by taking an apparent action awareness at face value is not reached by inference. It is no more inferential than is a perceptual judgement made by taking a perceptual experience at face value.

Because action awareness is not judgement or belief, a self-ascription of a mental action made by taking an action awareness at face value is a counterexample to the principle that knowledgeable mental self-ascriptions must be made by observation, by inference, or by nothing.[3] On the present view, that is a spurious trilemma. Action awareness is not perception, and can exist in the absence of perception of the action of which it is awareness. Judgements based on action awareness are not reached by inference, since action awareness is not judgement or belief. And judgements based on action awareness are not based on nothing, since action awareness is a real state of consciousness, available for rationalizing certain judgements. (These points apply equally to self-ascriptions of bodily actions too. The trilemma 'by observation, by inference, or by nothing' is similarly inapplicable to knowledge of one's bodily actions, when based on action awareness.) We should draw the conclusion that the model of observation is not the only model available for a substantive, non-inferential epistemology of first-person mental ascriptions.

As in other cases in which the content of a belief-independent state is taken at face value, there are (at least) two ways of developing an account of the matter. On one approach, the content of the belief-independent state does not involve conceptualization of the action-type or the content of the mental action. This is the option that holds that the content of the apparent action awareness is nonconceptual representational content. A theorist developing this treatment will be likely to hold that part of what it is to possess the concept of (say) judgement is to be willing apply the concept rationally to oneself in response to an apparent action awareness of one's judging a certain content. So a concept of an action-type, judgement, is individuated in part by its relations to action awareness of judgement; just as, on the perception side, an observational shape concept (say) is individuated in part by its relations to perceptual content with a certain nonconceptual representational content. On a different, conceptualist,

[3] The first sharp formulation of this view known to me is in P. Boghossian, 'Content and Self-Knowledge', *Philosophical Topics,* 17 (1989): 5–26, at p. 5.

approach, in the style of McDowell, it would be maintained that all personal-level content is conceptual, and this applies as much to the content of apparent action awareness as, on the conceptualist view, it applies to perceptual content.[4] On this view there would be no such thing as conscious action awareness without conceptualization of what it is awareness of.

The issues at stake in choosing between these two approaches are well-known from the case of perceptual content, and I will not pursue them at this point. As far as I can see, the thesis that we come to know of our mental actions by our action awareness of them is neutral between these opposing lines of thought. The thesis can consistently be accepted by the believer in nonconceptual content, and can consistently be accepted by his conceptualist opponent. In this area, there may seem to be a special problem for the friend of nonconceptual content. If the content of a judgement or decision, say, is conceptual, as it is, how can the action awareness of making the judgement or decision be nonconceptual? Here we must distinguish what the awareness is of at the level of reference, and how events, things, and properties at the level of reference are given in consciousness. A characterization of a state's content as nonconceptual has to do with how things are given, not which things are given. A state of consciousness can have a nonconceptual content concerning things that include concepts. This is something we should already recognize independently to be possible if we grant that there can be conscious thinking by children who do not have concepts of concepts and do not have concepts of intentional contents built up from concepts. It is one thing to be employing concepts, and have conscious states whose content involves those concepts. It is a further thing to conceptualize those intentional contents themselves.

Is action awareness philosophically explicable in terms that do not involve reference to subjective, conscious states and events? I call the claim that it is so explicable 'The Reducibility Thesis'. Under the Reducibility Thesis, however it is developed, action awareness is not something fundamental, and to understand the role of action awareness in our thought we must look to more fundamental conditions that do not involve consciousness. Any epistemological role played by action awareness would then be played by these more fundamental conditions not involving consciousness. But I dispute the Reducibility Thesis.

How might the Reducibility Thesis be developed? Can we say that action awareness consists in no more than an action's being a result of the operation of rational agency? That would need qualification on several fronts. (*a*) Making photocopies is an action of mine, but I need not have an action awareness that I am making copies. My action awareness is of pressing certain buttons on the machine. To accommodate this, the Reducibility Thesis could be confined to types of action that are basic for the agent, actions the agent does not, in the content of his intentions, do by doing something else. The defender of the

[4] J. McDowell, *Mind and World* (Cambridge, MA: Harvard University Press, 1994).

Reducibility Thesis would need to make this restriction to basic action-types both for bodily actions and for mental actions. (*b*) The Reducibility Thesis would also have to make some accommodation of what Brian O'Shaughnessy calls sub-intentional acts.[5] Tapping your toes, moving your tongue are actions. You can become aware of them, and indeed come to have a distinctive action awareness of them, but it is not clear that that action awareness was already there when the actions were first performed. The defender of the Reducibility Thesis may make various moves at this point. One would be to insist that there is action awareness even in these cases, but its content does not go even into short-term memory. Another would be to hold that the Reducibility Thesis holds only for the fully intentional acts of a rational agent. Both of these responses would need some work to become convincing; but let us leave speculation on how that might be done, because there is a deeper, and quite general, problem for the Reducibility Thesis.

It seems there could exist a being whose movements and whose changes in mental state are sensitive to the content of its beliefs and intentions, but whose tryings and actions, both bodily and mental—if actions they be—do not involve any action awareness, either real or apparent. These beings would have to perceive their bodily actions, through vision, touch, or proprioception, to know that they are occurring. Would such subjects be exercising rational agency as that notion is understood within the terms of the Reducibility Thesis? If so, then the notion of rational agency employed in the Reducibility Thesis is so thin that it seems incapable of capturing action awareness at all. But if such subjects are not so conceived of possessing rational agency, it seems that action awareness, of both bodily and mental actions, has to be conceived as a co-ordinate element in rational agency in its own right. An explanation of the epistemology of action, both bodily and mental, has to go beyond materials that could equally be present in cases that are wholly non-conscious on the action side.

This is a point that bears not only on the Reducibility Thesis, but equally on any attempt to explain certain kinds of self-knowledge in terms of agency alone. We need to recognize a co-ordinate, and irreducible, element of consciousness in rational agency and action awareness as we actually have it. In my judgement, many illuminating recent discussions of self-knowledge and agency work only because there is a background assumption that we have action awareness of our bodily and mental actions.

A view that is distinct from the Reducibility Thesis, but which still moves in a deflationary direction, is that though there is such a thing as action awareness, it is really no more than is involved in the wider species of non-inferential knowledge. A person can know that Beethoven was born in Bonn, and this knowledge need not be based on anything more than the propositional impression, delivered by

[5] B. O'Shaughnessy, *The Will* (Cambridge: Cambridge University Press, 1980), ii, ch. 10, 'The Sub-Intentional Act'.

memory, that Beethoven was born in Bonn. Is action awareness any different from the mere propositional impressions involved in such non-inferential knowledge? In fact the mere propositional impression that one is acting in a particular way, whatever its source, is something weaker than action awareness, and is not sufficient for the distinctive phenomenology of action awareness. One symptom of this difference is that action awareness makes available a distinctive variety of demonstrative ways of thinking of actions given in one's action awareness. You have an action awareness of *this* raising of your arm. There would not even exist a distinctive demonstrative *this raising* unless you had this action awareness. If action awarenesses, real or apparent, were mere propositional impressions, it would be unintelligible how the conceptual component *this raising* could exist to be available for use in your thought, for no account would be available of how its reference is determined, in the absence of real or apparent action awareness. Though action awareness is not perceptual awareness, the problems for such a position are structurally entirely analogous to those attending a position that aims to reduce perceptual experiences to mere propositional impressions. Those propositional impressions, in the perceptual case, would have to contain perceptual demonstratives such as *that cup, that door*, if the account is to get off the ground at all. But just as in the action awareness case, these demonstratives are individuated by their relations to perceptual experience. Mere propositional impressions, both in the action and in the perception cases, are inadequate to a description of these phenomena.

2. WHAT MAKES THE WAY IDENTIFIED IN THE ACTION AWARENESS ACCOUNT A WAY OF COMING TO *KNOW*?

It is widely accepted that there is a range of observational concepts—concepts of shape, size, orientation, colour, texture, amongst others—that are individuated in part by the fact that certain perceptual experiences give reasons to apply these concepts to objects or events presented in those perceptual experiences.[6] Theorists differ on how this individuation works, but there is less disagreement that there is some such individuative link between these concepts and perceptual states. What makes such an individuative link possible is in part the existence of perceptual states with representational content. I suggest that the representational content of action awareness provides a similar resource for the individuation of certain concepts of mental action. Some concepts are individuated in part by the fact that action awareness gives reason to apply these concepts.

[6] For further discussion, see *A Study of Concepts* (Cambridge, MA: MIT Press, 1992), ch. 3, and 'Does Perceptual Experience Have a Nonconceptual Content?', *Journal of Philosophy,* 98 (2001): 239–64.

One clause in a formulation of the possession condition for the concept *judging that p* should treat the case of first-person application in the present tense. It should state that, in the absence of good reasons for doubt, an apparent action awareness of his judging a given content gives reason for a thinker to accept *I judge that p*. Here, the action awareness in question has a content to the effect that the thinker is himself judging that *p*. When the action awareness is awareness of a judgement, and a thinker self-ascribes in accordance with this possession condition for first-person ascriptions of judgements, his self-ascriptions are sensitive to the event's being a judgement. Quite generally, making a judgement in accordance with one of the clauses of a possession-condition for a concept in the content of the judgement is a way of coming to know the content of the judgement in question.[7] The action-blind subjects considered in the response to Question 1 could not, incidentally, exercise this concept of judgement in making ascriptions to themselves (if indeed they could possess concepts at all), since they lack action awareness of their judgements.

An account of possession of these mental-action concepts must also have a clause dealing with third-person ascriptions. To understand third-person ascriptions of these concepts is to have tacit knowledge that their correctness requires the subject of the attribution to be in the same state the thinker is in himself, when a first-person attribution is correct. Under this approach, the bridge from first-person ascriptions to third-person ascriptions is built using tacit knowledge involving grasp of an identity relation. This is a structural feature it shares with certain accounts of the bridge from observation-based to non-observation-based applications of an observational concept, and with accounts of the bridge from first-person ascriptions of other conscious states to other-person ascriptions.[8] Theorists may, however, present other, competing accounts of the bridge, consistently with accepting the first-person clause I have been offering. I will not pursue this further here, since our main focus is on the first-person clause.

Does the first-person clause I have advocated embody a perceptual model of the self-ascription of certain attitudes? It does not. Action awareness is not perceptual awareness; a subject can have action awareness of something without having any perceptual awareness of it. It is no consequence of the present view that, when judging in accordance with the relevant possession conditions, one perceives or observes one's judgements or decisions. Nor does the present view postulate intermediaries which would somehow be an obstacle to knowledge of one's own judgements, decisions, and other mental actions.

[7] A principle I proposed and argued for in *A Study of Concepts* (Cambridge, MA: MIT Press, 1992), 157–8, and in *Being Known* (Oxford: Oxford University Press, 1999), ch. 2.

[8] See C. Peacocke, *The Realm of Reason* (Oxford: Oxford University Press, 2004), 48–9; '"Another I": Representing Conscious States, Perception and Others', in J. Bermúdez (ed.), *Thought, Reference and Experience: Themes from the Philosophy of Gareth Evans* (Oxford: Oxford University Press, 2005); 'Justification, Realism and the Past', *Mind,* 114 (2005): 639–70; and my book *Truly Understood* (Oxford: Oxford University Press, 2008), ch. 5.

In the case of genuine perception of material objects and events, one would insist that a subject perceives an object or an event itself in a certain way. Far from perception inserting an intermediary that prevents access to the material objects and events themselves, it is perception that makes possible such access to the events and objects themselves. The same is true of action awareness. We should take the grammar at face value. In the bodily case, the subject is aware of his action itself, his clenching his fist, say, and he is aware of it as his clenching his fist. It is as wrong to think of action awareness as some epistemically problematic intermediary preventing access to the events and objects themselves as it is wrong to think of perceptual experience as an epistemically problematic intermediary between subjects and the world.

Action awareness that one is φ-ing is a factive notion. It implies that one is φ-ing (arguably it also implies that one knows one is φ-ing).[9] As some of the earlier examples show, there is such a state as mere apparent awareness that one is φ-ing, a state whose content can be false. Someone might argue that all we, as agents, ever have is mere apparent awareness that we are φ-ing. This is a cousin of the argument from illusion in perception, applied here on the side of action.

The argument in the action case is no more sound than its perceptual cousin. When an apparent action awareness that you are φ-ing stands in the right complex of relations to your φ-ing, the apparent action awareness *is* genuine awareness that you are φ-ing. The complex of relations in question is different from those involved in the perceptual case. The relations in question run predominantly from the mind to the world in the action case, rather than the opposite direction of the perceptual case. But the fallacy involved in the argument from illusion is the same in both the perception and the action cases.

Even if the treatment I am offering is not vulnerable to the argument from illusion, it may be thought that it is still open to the objections McDowell has raised against what he calls 'hybrid' accounts of knowledge.[10] As applied to the present subject matter, the complaint would be that, on the offered account, there could be a pair of cases in both of which the subject has the apparent action awareness that entitles him to self-ascribe a mental action, yet in one of these cases the self-ascription is true, and the other is false. The objection, to summarize it, is that if this is possible, the self-ascription cannot amount to knowledge in the first case. This is not a chapter about

[9] I use the notation 'φ-ing' to formulate these generalizations, but this should not be taken to imply that it is only continuing events of which one can have action awareness. One can have an action awareness of something that does not take time, both in the bodily and in the mental domains. Stopping talking can be an action, and the agent can have an action awareness of it. It is not a continuing event. Judging and deciding are also not temporally extended processes, but the subject can have action awareness of them too.

[10] 'Knowledge and the Internal', repr. in his collection *Meaning, Knowledge and Reality* (Cambridge, MA: Harvard University Press, 1998).

general epistemology, so I will not divert the discussion into what would need to be an extended consideration of the status of hybrid theories. The main message of this chapter is the role of action awareness in the knowledgeable self-ascription of mental actions. That message can certainly be incorporated into a McDowellian epistemology if one so wishes. That incorporation would proceed by first insisting that in genuine action awareness that one is φ-ing, the subject's mind is embracing the fact that he is φ-ing. The position would then go on to say that the subject, in judging that he is φ-ing, is simply taking this factive state at face value, is endorsing its representational content. Action awareness that one is φ-ing would, on this McDowellian incorporation of the point of this chapter, play the same epistemic role in relation to certain self-ascriptions of actions as perceptual awareness that p plays, on his account, in attaining perceptual knowledge that p. On the McDowellian approach, in the case in which subject has a mere apparent action awareness, the kind of state which gives his reason for making his self-ascription of φ-ing is not the same as the kind in which it is genuine awareness of his φ-ing. So the alleged objections to hybrid theories would not get a grip. I am not endorsing this McDowellian approach. The issues involved in assessing it are orthogonal to the main thesis of this chapter. My point is just that the idea that action awareness of our mental events is important for the epistemology of some mental self-ascriptions can be acknowledged on both McDowellian and non-McDowellian positions in general epistemology.

Although action awareness is distinct from perceptual awareness, there is a significant parallelism of abstract structure in the perception and action cases on the view I am advocating. There is a structure of rational entitlement in which the entitling state has representational content; and one can be mistaken about whether the content of the entitling state is correct (or whether it is really a factive state that one is in—the parallelism is equally present on a McDowellian treatment). There is a danger here that we may endorse the following fallacious argument:

> Mental actions are not given to their subject under a perceptual mode of presentation.
>
> Hence,
>
> One possible source of error is absent for mental actions that is present for perceptual beliefs about the external world; that is, self-ascriptions of mental actions have a certain domain of infallibility that perceptual beliefs do not.

The premise of the argument just displayed is true. What follows the 'that is,' in the conclusion is false. An apparent action awareness can have a false content, just as an apparently perceptual experience can have a false content. (In the case of mental action, this is the ground of the possibility of one form of self-deception: it may seem to one that one is forming a belief when in fact

one is not.) The fact that action awareness is not perceptual-awareness does not give it any kind of infallibility, however limited, that perceptual awareness lacks. The premise of the fallacious argument rightly alludes to the distinction between action awareness and perceptual awareness. This difference in kind does not by itself produce any kind of philosophically significant restriction on fallibility. If there are restrictions, their sources lie elsewhere.[11]

The modest amount I have said so far about mental actions and concepts of them fits a broadly rationalist model of entitlement. The possession-condition for concepts of mental actions contains a clause about first-person present-tense ascription that says that the thinker has reason for making such ascriptions in the presence of suitable apparent action awareness. This accords with a general model under which an entitlement to make a transition to a given judgement always has some a priori component that is founded in the nature of the contents involved in the judgement and the reasons for it, and in the nature of the mental states involved in the transition. Here the relevant a priori component is found in a transition (it is a form of the relatively a priori). A thinker is entitled to take the content of an event of apparent action awareness at face value, in the absence of reasons for doubt. The claim of the existence of some a priori component in every entitlement was the general position I defended in the early chapters of *The Realm of Reason*.

We need, however, to have a much better understanding of how exactly apparent action awareness provides a thinker with entitlement to make judgements about his own actions. The understanding we seek should explain how relying on apparent awareness furthers the goal of making judgements that are true.

In the second chapter of *The Realm of Reason*, I distinguished three levels at which one can characterize the entitlement relation. There is, first, the level of instances of the relation. There is, next, a second level of true generalizations about the relation, generalizations that have as instances truths at the first level of characterization. At a third level are principles which explain why those generalizations at the second level are true (and thereby also explain the instances). The third level, as the explanatory level, is the one we should seek to elaborate further in the case of action awareness and the self-ascription of mental actions.

What makes an apparent action awareness one of clenching one's fist, or raising one's arm, or judging or deciding some particular thing, is that, when these and the subject's other mental states are properly connected to the world, they are caused by events (tryings) that cause a clenching of the first, a raising

[11] I may have been guilty of the fallacy identified in this paragraph. There is a whiff of it in my contribution to a symposium with Tyler Burge on self-knowledge. See my 'Entitlement, Self-Knowledge and Conceptual Redeployment', *Proceedings of the Aristotelian Society,* 96 (1996): 117–58, esp. p. 126: 'brute error is impossible. It is impossible precisely because, in these psychological self-ascriptions, there is nothing that plays the role that experience plays in genuine observational knowledge of physical objects.'

of one's arm, or a judging or deciding of some particular content. That is, the mental states of apparent action awareness are relationally, and in a certain sense externally, individuated. What makes them the states they are is the fact that when all is functioning properly, and the states are properly embedded in relation to the subject's other mental states, his body, and the external world, they have a cause which also causes what they are as of—what they represent as being correct. My own view is that the easiest way for such complex, relationally individuated states to occur is for states of their kind to have evolved by a selection process, one which favours the occurrence of those states whose representational content is correct. In taking apparent action awareness at face value, one is judging that things have come about in what is in fact the easiest way for them to come about.

Under this approach, once again it appears that, although action awareness is distinct from perceptual awareness, the structure and underlying explanation of entitlement relations involved in relying on action awareness is arguably the same as that underlying perceptual entitlement. The outline just given of why there is an entitlement to take certain action awarenesses at face value is entirely parallel to an argument that there is an entitlement to take certain observational contents of apparent perceptual experience at face value.

This outline of how action awareness entitles a thinker to make self-ascriptions of bodily and mental actions is given for the neo-rationalist approach to entitlement that I myself favour. That approach is opposed to purely reliabilist accounts of entitlement that do not include rationality requirements that are distinct from considerations of reliability. But I should note that pure reliabilists, and no doubt reliabilists of other stripes, could equally accept the importance of taking apparent action awareness at face value in the account of how we come to know our own mental actions. Taking apparent action awareness at face value is not at all something proprietary to neo-rationalists; it can serve many other comers too.

3. IS THE ACTION AWARENESS MODEL OPEN TO THE SAME OBJECTIONS AS PERCEPTUAL MODELS OF INTROSPECTION AND AWARENESS?

It is widely held amongst current philosophers of mind that models of introspection that treat it as a form of perception are untenable. I have repeatedly emphasized that action awareness is not perceptual awareness. But action awareness, as a source of self-knowledge, does involve a conscious state that stands in complex causal relations to what it is an awareness of. Action awareness is also, as I have equally emphasized, to be sharply distinguished from judgement that one is performing a certain action; and is also to be distinguished from awareness

merely of trying to perform the action. So there is a pressing question: do the objections to perceptual models of introspection, suitably adapted, apply equally to action awareness models of first-person knowledge of mental action?

One of the most interesting and general arguments against perceptual models of introspective knowledge has been developed by Sydney Shoemaker in the second of his Royce Lectures, 'Self-Knowledge and "Inner Sense" ', in the lecture titled 'The Broad Perceptual Model'.[12] His discussion of the perceptual model of introspection includes the following theses:

> Shoemaker's Thesis (1): Under the perceptual model, 'the existence of these [perceptually known] states and events is independent of their being known in this way, and even of there existing the mechanisms that make such knowledge possible' (pp. 224–5).

Shoemaker calls his Thesis (1) 'the Independence Condition' (my capitals). I agree that the Independence Condition must be a commitment of any conception worthy of being called perceptual.

> Shoemaker's Thesis (2): The Independence Condition implies the possibility of what Shoemaker calls 'self-blindness'.

To be self-blind with respect to certain mental facts or phenomena is to be able to conceive of them—'just as the person who is literally blind will be able to conceive of those states of affairs she is unable to learn about visually' (p. 226)—but not to have introspective access to them. The possibility of such self-blindness, Shoemaker writes, 'I take to be a consequence of the independence condition that is built into the broad perceptual model of self-knowledge' (p. 226). I call his Thesis (2) 'the Thesis of the Independence/Self-Blindness Link', or 'the Link Thesis' for short.

> Shoemaker's Thesis (3): Self-blindness is not a genuine possibility in respect of pains; nor in respect of perceptual experience; nor in respect of the will and intentional action; nor in respect of beliefs (sections II, III, IV and V of the Second Lecture respectively).

He elaborates: 'it is of the essence of many kinds of mental states and phenomena to reveal themselves to introspection' (p. 242). It follows from Shoemaker's Thesis (3), together with Thesis (2), that the Independence Condition is false for pains, perceptual experience, the will, intentional action, and beliefs. It also follows in turn by modus tollens from Thesis (1) that the perceptual model of introspection is false.

> Shoemaker's Thesis (4): The correct account of the relation of these mental events and states (pain, experiences, intentions, actions, beliefs) to awareness

[12] Repr. in his collection *The First-Person Perspective and Other Essays* (Cambridge: Cambridge University Press, 1996), ch. 11.

> of them needs to draw on the distinction between the core realization of a state and its total realization (pp. 242–3).

The core realization comes and goes as the mental state comes and goes. 'The total realization will be the core realization plus those relatively permanent features of the organism, features of the way its brain is "wired", which enable the core realization to play [the causal role associated with that state]' (pp. 242–3). Adding rationality, intelligence, and possession of the concept of the concept of belief to a first-order belief enables the core realization of the first-order belief to play a more encompassing role. When this surrounding material is present, a first-order belief and the second-order belief that one has that belief have the same core realization. The total realization of the first-order belief is a proper part of the total realization of the self-ascriptive belief that one has the first-order belief (243).

If Shoemaker's arguments in his Theses (1) through (4) are sound, their applicability is not restricted to the perceptual model of introspection. They apply to any subject-matter for which the Independence Condition is fulfilled, and for which self-blindness is not a possibility. This generalizability of Shoemaker's argument is part of its interest and challenge.

It certainly appears that, if the argument is sound, it must generalize to apply against the action awareness account of our knowledge of our own actions (bodily or mental). Action awareness of a particular action is certainly distinct from the action itself. The real or apparent action awareness lies on a different causal pathway from the action itself. The awareness is caused by an initial trying, or some initiating event, which trying or event also causes the effects (the arm's rising) that are required for there to be an action of the kind in question. Even if there is an argument that tryings must, at least in central cases, involve awareness of those tryings, the trying and the awareness of trying is distinct from action awareness. The relation between some constitutive components of the action and the action awareness of the action is causal. It does not seem to be an option to say that there are no causal-explanatory elements at all in the action awareness account.

But this then seems to leave it at least metaphysically possible that there be actions without the distinctive kind of action awareness that we enjoy. This is precisely the case of action-blindness we considered in addressing Question 1 above. What makes an event in that envisaged world an action is the fulfilment of the same condition as makes something an action in the actual world: it is caused in the right kind of way by a trying. The actions in this non-actual world would be explained by their agents' contentful intentional states (conceptual or nonconceptual). To fail to acknowledge a category of actions in this possible world would be to miss an explanatorily significant category of events. Action-blind subjects would have to know about even their own bodily and mental actions in ways in ways in which they learn about other events.

Even when the formation of one of their beliefs is explained by their other mental states, in wholly intelligible ways, they have no distinctive awareness that they have formed that belief. These subjects would in some respects be as opaque to themselves as another person may be to them. Still, it should not be denied that some bodily events and some mental events in this imagined world are actions, and are appropriately explained by the subject's mental states. There is a plausible case to be made that there can be actions without action awareness, and if this is right, then Shoemaker's Independence Condition is met for actions.

So if Shoemaker's argument is sound, it would follow that the action awareness account is committed to the possibility of self-blindness in respect of such mental actions as judgements, decisions, and the rest. That is what his Thesis (2), the Thesis of the Independence/Self-Blindness Link implies. It is this Link Thesis on which we need to focus in assessing the bearing of Shoemaker's argument on the action awareness account of knowledge of our mental actions.

Whenever something is impossible, one should ask: what is the explanation of the impossibility? If self-blindness is not possible in respect of certain states and events, it may be that the explanation of the impossibility traces to the conditions required for possessing concepts of those states and events, rather than being explained by the failure of the Independence Condition. Actually it seems to me that further reflection on Shoemaker's own initial illustration of a genuine case of self-blindness, of the genuinely blind person who is able to conceive of the states of affairs that she cannot see to obtain, supports this alternative explanation. The blind person can conceive of objective states of affairs involving objects, events, their properties, and spatial relations only because she is capable of perceiving these things and properties in at least some other sense modality—by touch and hearing, for instance (or else because she was once able to see, and knows what it would be to have visual experience of objective states of affairs). If we are asked to entertain the possibility of someone who is supposed to have the conception of material, spatial objects, and events, whilst also lacking all such perceptual faculties, and lacking all knowledge of what it would be like to have them, it seems reasonable to question whether this is a genuine possibility. It is such faculties that make possible the thinker's possession of concepts of objects and events that may be perceived in one or more sense modalities. If this is so, then there could not be someone who is capable of no perceptual states at all, yet has the concept of objects and events he cannot perceive. The explanation of this impossibility has, however, nothing to do with failure of the Independence Condition. The Independence Condition holds as strongly as ever for conditions concerning external objects, events, and many of their properties and relations. It would be quite wrong to move from the impossibility of someone who both lacks all perceptual faculties and who conceives of objects and events he cannot perceive to the conclusion that the existence of material objects and events is not independent of our ability to conceive of them, to

perceive them, or to know of them. Their existence is so independent, in all these respects.

Structurally, the position here is as follows. The claim of the possibility of self-blindness with respect to some states of affairs is a claim of the form $\Diamond$ (p & $\sim$q): it's possible that the subject has the concept of those states of affairs and yet does not have a certain kind of access to them. When self-blindness is not possible, we have something of the form $\sim\Diamond$ (p & $\sim$q) holding. A proposition of that last form is equivalent to the corresponding proposition of the form $\Box$ (p $\rightarrow$ q). The explanation of this necessity's holding may simply be that, necessarily, whenever the conditions for the subject's possessing concepts of those states of affairs hold, the subject also has a certain kind of access to them. Such access may be involved in the possession conditions for the concepts in question. This can all be true consistently with the Independence Condition's still holding for the states of affairs in question.

A case which seems to me clearly to exemplify this possibility is that of pains and beliefs about pains. An animal can have real pains (not just some surrogate or proto-pains), without having the concept of pain, and hence without having any ability to think about its pains as pains. The existence of pains is independent of their being known about, as the Independence Condition requires. Self-blindness is nevertheless arguably impossible for the state of being in pain. The explanation for this is the widely accepted point that part of what is involved in having the concept of pain is a willingness to judge, and judge knowledgeably, that one is in pain when one is in pain, where the pain itself makes rational the thinker's judgement. This explanation does indeed not have anything to do with failure of the Independence Condition.

These points also highlight the fact that the sense in which pain is something essentially open to introspection—a consideration Shoemaker uses in the intuitive defence of his Thesis (3)—is to be distinguished from the claim that its nature is constitutively dependent on what its possessor would judge about it in specified circumstances. Introspection is a matter of the occupation and direction of attention, rather than something to be characterized at the level of judgement.

This consideration of the case of pain shows two things:

(*a*) There are relatively uncontroversial instances in which we have the Independence Condition holding, consistently with the impossibility of self-blindness. It follows that we cannot take the failure of the Independence Condition as the explanation of an impossibility of self-blindness. Shoemaker writes of introspection, contrasting it with perception, that 'the reality known and the faculty for knowing it are, as it were, made for each other—neither could be what it is without the other' (p. 245). We are committed to disagreeing with this in one direction: pain could be what it is independently of the presence of the capacity for, and the nature of, thought about pain. The concept of

pain is, however, certainly made for knowing about pains. The explanation of the impossibility of self-blindness in the case of pain has more to do with the nature of the concept *pain* than with the nature of pain. It would be wrong, however, to say that the explanation has nothing to do with the nature of pain itself. It is because pains are conscious, subjective events that pain itself is capable of featuring in the possession-condition for the concept *pain*.

(*b*) The second lesson is that if, as is also widely accepted, we do not perceive our pains but simply experience them, the Independence Condition can hold even in a case in which the perceptual model itself fails.

An explanation of the impossibility of self-blindness in the case of one's own mental actions is analogous in some respects to that just given for the case of pain, and is disanalogous in others. The explanation is partially analogous in respect of the role played by the possession conditions for such concepts as those of judgement and decision. To possess the concept of judgement involves applying it to oneself in response to one's action awareness of one's own judgements. If a thinker is capable of doing this, he will not be self-blind in respect of his mental actions. His ability to conceive of judgements, decisions, and other mental actions as such is constitutively dependent upon his ability to come to know of them in certain way. (A thinker might lose the ability to have states of action awareness of his mental actions, or actions of given type, just as someone may become blind, or blind to certain types of states of affairs. Provided the thinker still knows what it is to be in such states, the corresponding concepts of the states are still available to him.)

It would be an objection to this account of the nature and limits of the impossibility of self-blindness in the case of mental actions if there were a different account of possession of the concepts of judgement, decision, and other mental action-types, an account that does not give an essential, constitutive role to action awareness. I do not know how such an account might run. Could an alternative account talk of the thinker's tacit knowledge of an individuating role for judgement, or decision, or some other action-type, in a psychological economy? Such tacit knowledge seems unnecessary in simply making a knowledgeable present-tense self-ascription of an action in rational response to an action awareness of one's performing such an action. For third-person (or other-tense) ascriptions, once one has the role of action awareness in the first-person, present-tense case, a thinker's understanding of the other cases can consist simply in his tacit knowledge that they are correct if their subject is in the same state as someone who is genuinely action-aware of his performance of the action-type in question. In my judgement, this description of the tacit knowledge is more faithful to what has to be explained than attribution of tacit knowledge of a quite specific psychological role for the action-type in question. Insofar as ordinary thinkers are able to reach conclusions about the role of a

mental action-type in a thinker's psychology, it is by way of application of this identity-condition.

A major respect in which the cases of pain and action awareness are disanalogous is that in making a self-ascription on the basis of action awareness, a subject is endorsing the content of a representational state. Pain is not, in my view, a representational state (or at least, it is not necessary for the purposes of this account that it be so, unlike the case of action awareness). This difference means that we need an account that addresses the question of why we are entitled to take the representational content in question at face value, as touched upon in the preceding section. Once again, though action awareness is not perceptual awareness, the need for such an account is something shared with the case in which a perceptual experience is legitimately taken at face value.

To summarize this critique to this point: (*a*) there are counterexamples to Shoemaker's thesis that the Independence Condition implies the possibility of self-blindness; (*b*) there are alternative explanations of the impossibility of self-blindness, to the extent that it is impossible, consistently with rejection of the perceptual model of introspection; and (*c*) the explanation of the impossibility of self-blindness has more to do with the nature of the concepts involved in thought about these mental states and events, than in the nature of the events themselves.

What, however, of Shoemaker's own positive explanation of the impossibility of self-blindness in the cases he discusses? There is some reason to doubt that the distinction between the core and the total realization of a state, and Shoemaker's proposal about its extension in cases of introspective knowledge, can do quite the work he requires of it. Shoemaker's view is that the core realization—the realizing state that comes and goes as what it realizes comes and goes—is the same for the mental state thought about and the self-ascription of the state. But since a judgement that one is in a certain kind of mental state requires employment of one's concept of that state (and of oneself, and of the present), the structured state that realizes this judgement is much more plausibly identified as something causally downstream from the mental state that verifies the content of the judgement about oneself as correct. The realizing state must have sufficient structure for it to realize judgement of a structured intentional content, involving concepts combined in a quite specific way. I will not pursue this further here, partly from limitations of space, and partly because the issues are not specific to issues of awareness and self-knowledge. I just note that this second objection will be compelling to those who see something in the arguments, marshalled some years ago in debates about the language of thought, to the effect that the causal-explanatory powers of states with intentional content require corresponding structure in their realizing states.[13] While it is true that Shoemaker includes in his

[13] J. Fodor, *The Language of Thought* (New York: Thomas Crowell, 1975); 'Appendix: Why There Still Has to Be a Language of Thought', in J. Fodor, *Psychosemantics: The Problem of*

total realization whatever it is that realizes possession of particular concepts, that would not be enough to meet the concerns of these critics. When someone makes the judgement that he is in pain, it is not merely that he possesses the concept of pain. The state that realizes his judgement must also realize the activation or use of his concept of pain, and thus be ready for inferential interactions involving the concept in other premises. Simply being in pain seems to fall short of that. The distinction becomes vivid when for, instance, one thinks one is in pain when the dentist approaches with some terrifying instrument. The dentist then says 'I haven't even touched you yet!' In the patient's rush to judgement, he judges that he's in pain, and the realization of this will involve the activation of concepts, and the placing of symbols for the concepts, suitably combined, into the 'belief-box' on theories endorsing the existence of a language of thought. But the subject is not really in pain (nor is a possession-condition relating possession of the concept of pain to the occurrence of pain undermined by such impulsive cases). The most natural treatment of such examples is to say that, even for core realizations, the core realization of pain is distinct from the core realization of the judgement that one is in pain.

4. HOW DOES THE ACTION AWARENESS ACCOUNT RECONCILE EXTERNALISM ABOUT INTENTIONAL CONTENT WITH PRIVILEGED SELF-KNOWLEDGE?

Twenty years have now passed since intensive discussion began about the reconcilability of distinctive self-knowledge with the external character of intentional content. The first decade of this discussion has by no means quieted the objections of those who say that some theories of self-knowledge make such reconciliation impossible, so that we must either abandon those theories, or reject externalism about content. Is the action awareness account of our knowledge of our own mental actions an account that makes such reconciliation impossible? And if it does not, what is its positive account of the nature of the reconciliation?

Doubts about the possibility of reconciliation on certain models of self-knowledge have been concisely articulated, and endorsed, by Crispin Wright, writing about halfway (1996) through this twenty-year period. Wright considers the model of self-knowledge as inner observation, and writes:

> I want to say that . . . in the sense in which an image or mental picture can come before the mind, its intentionality cannot. . . . Both a sunburned arm and a triangle can be

Meaning in the Philosophy of Mind (Cambridge, MA: MIT Press, 1987); M. Davies, 'Concepts, Connectionism and the Language of Thought', in D. Rumelhart, W. Ramsey and S. Stich (eds.), *Philosophy and Connectionist Theory* (Hillsdale, NJ: Erlbaum, 1991).

presented as ordinary objects of observation, and each sustains, *qua* presented under those particular respective concepts, certain internal relations: the sunburned arm to the causes of its being in that condition, and the triangle to, for instance, other particular triangles. And the point is simply that while the identification of the triangle as such can proceed in innocence of its internal relations of the latter kind . . . recognition of the sunburned arm as just that cannot proceed in like innocence but demands knowledge that its actual causation is as is appropriate to that mode of presentation of it.[14]

Wright attributes to Wittgenstein, and finds convincing, the point that 'the internal relations to the outer, of whatever sort, are all of the latter—sunburn-style—kind; and hence there is indeed a standing puzzle in the idea that an appropriate characterization of them, incorporating such intentionality, is somehow vouchsafed to their subject by something akin to pure observation' (p. 343).

Does Wright's objection apply equally to the action awareness account? In the perceptual case, as Wright implies, there is a partition between properties such as that of being sunburned that cannot be known to be instantiated simply by taking perception at face value, and those such as shape, colour, orientation, surface texture, and so forth, which can be known to be instantiated simply by taking perceptual experience at face value. (A state's representing these latter properties is also a matter of its external relations, incidentally, a fact that should give us pause about the direction in which the argument is going.) An analogous partition of properties, as thought about in given ways, can be made for action awareness. I cannot, from action awareness alone, come to know that the copying machine whose lid I am closing was manufactured in Taiwan. Knowing that requires knowledge of its history that is not given in action awareness. But action awareness can make available knowledge that I am closing the machine's lid, at a certain speed, with a certain force, and that I am doing it now. So, in the case of mental actions, the crucial question to address is this: is the intentional content of a mental event or state to be grouped with the property of being made in Taiwan, or with the properties which you can know about simply from your action awareness?

The intuitive, pre-theoretical answer to this question is that we have an action awareness of the full intentional content of our judgements, decisions, and other mental actions. We are aware that we are judging that New York is hot in the summer; we are aware that we are deciding to spend the summer in a cooler place. A judgement may also be a manifestation of a neurosis, may be an unconscious excuse for not staying in New York, or many other things that are to be grouped with the machine's being made in Taiwan. But the intentional content of the judgement, decision, or whatever mental action-type is in question does seem to be so available. What is the explanation of this fact? Whatever the

[14] C. Wright, *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations* (Cambridge, MA: Harvard University Press, 2001), 342–3.

explanation, it will have to have a certain generality. When we know what we are judging or deciding, on the basis of action awareness, we know the content of our judgement or decision, whatever its conceptual constituents. You can have an action awareness of your judging that *p*, whatever the content *p* may be, whether the conceptual content *p* is observational, theoretical, moral, or anything else.

When you judge, on the basis of an action awareness, that you judge that New York is hot in the summer, you are thinking of yourself as having an attitude that involves a certain concept, that of being hot. You think of that concept in a particular way. There are many different ways of thinking of a concept (as there are of thinking of anything at all): but in this second-order judgement about your first-order attitude, you are thinking of the concept *hot* under its canonical concept. The canonical concept of a concept F, can(F), is something made available by the concept F itself. There are various ways of elaborating this kind of availability in more detail, the various ways corresponding to various theories of how concepts are individuated. Suppose, as a starting point, we think that a concept is individuated by what has to be tacitly known about the condition for something to fall under that concept. Then we can state what is distinctive about the canonical concept of a concept thus:

(*) For an arbitrary concept C to fall under the canonical concept of the concept F (under can(F)) is for C to be such that: the fundamental condition for something to fall under C is the same as the fundamental condition for an object to fall under F.

We can call this a *leverage* account: it leverages an account of the condition for something to be F into an account of a particular higher-level concept of that concept. I am taking it that a thinker could not have the concept F itself (as opposed to some mode of presentation of F) in the content tacitly known in knowing what (*) states unless he has enough of a grasp of the reference-condition for F to be attributed with attitudes containing F in their content.

Under the leverage approach, the canonical concept of a concept F is unique. Under the model of the tacit knowledge of reference-conditions, there is only one canonical concept of the concept F built according to the pattern of (*). Clearly also only concepts can have such leverage-involving canonical concepts of themselves, since the very condition for something to be in the extension of the canonical concept of F requires it to be a concept. For something x that is not a concept, *the canonical concept of x* is something that is not, along these lines at least, well-defined.

This leverage account has two consequences.

(i) Reasons, in given circumstances, for accepting or rejecting particular contents containing the concept F become reasons for accepting or rejecting corresponding contents containing the canonical concept of F. If something

is a reason for judging Fa, in given circumstances, then it is equally a reason for judging that a falls under the concept thought of as the canonical concept of F. Since the fundamental condition for a concept to fall under the canonical concept of F is that it has the same satisfaction-condition as F, any reason for judging that something is F will equally be a reason for judging that it falls under the concept picked out by the canonical concept of F.

(ii) Any externalist features in the individuation of the concept F will be inherited by the concept *the canonical concept of F*.

We can now focus on the transition from a thinker's having

an action awareness of his judging that New York is hot in the summer

to his judgement of this intentional content, where the senses referred to are thought about under their canonical senses:

<Judge>^<I>^(can(<hot>)^can(<New York>)^can(<in the summer >)).

Here '<A>' denotes the sense expressed by A (I omit other formalities). This transition from the action awareness to the judgement is a priori valid. In any context in which the thinker has a genuine, and not merely apparent, action awareness of judging that New York is hot in the summer, it will also be true that he judges that it's the concept *hot* that he judges New York to fall under in the summer. The same applies to the canonical concept of any other concept F in place of the concept *hot*, however externally or historically individuated the concept F may be. The reason-giving state in this transition is as externally individuated as the content of the judgement that it rationalizes.

This transition in thought from action awareness to judgement is totally different from the transition, unwarranted without further information, from

a perception of a reddish arm

to a judgement

this arm is sunburned.

This latter transition does, just as Wright says, need further information about the causes of the redness on the arm, in a way the preceding transition does not need any further information for its legitimacy. Unlike the transition to a self-ascription of a judgement based on action awareness, the content of the judgement about sunburn involves external factors, about the cause of the redness, whose presence is not ensured by veridical perception of a reddish arm.

There may be a sense of unease about this reconciliation of externalism and the distinctive knowledge of mental actions, a sense that there is some kind of

cheating going on. This sense may stem from the thought that the treatment given here is like that of someone who insists, correctly, that the recognitional concept of water is externally individuated, and that we know our thoughts are water-thoughts. There is a clear sense in which one can possess a recognitional concept of water without knowing which liquid it is, in the sense of not knowing its chemical composition. Does a similar objection apply against the account I have offered of action awareness of the conceptual contents one is judging? Is the account consistent with the thinker's not knowing which concepts are in question?

I reply that because one is employing the canonical concept of a concept in making judgements about the contents of one's thoughts, one does, by contrast with the chemical characterization of water, know which concept is in question. It is precisely the force of the leverage accounts to make it clear that any such seeming gap is really closed. Under the leverage accounts, you know as much about which concept is in question when you think of it as *the concept F* as there is to know. All the conditions that contribute to the individuation of the concept F itself contribute to the individuation of the canonical concept of F. From the leverage accounts, as noted, reasons for making first-order judgements containing the concept F are transmitted to reasons for making suitably corresponding judgements containing the higher-order concept *the concept F*. Any requirements on knowing which concept that are met when one is simply using the concept F will equally be met when one uses the canonical concept of the concept F. It is always an answer to the query 'Which concept is in question?' to say 'It's the concept F', where this answer employs the canonical concept of F, rather than some descriptive mode such as 'the concept discussed in Chapter 6 of this book'.

There will, for any given concept, be empirical conditions met by a given thinker who employs that concept, conditions not extractible simply from the nature of the concept itself. They will include such matters as the nature of that particular thinker's mental representations underlying his possession of the concept, and the particular computational procedures involving it that he employs. There is manifestly an important area of study that consists in the empirical investigation of these empirical matters involving concepts as possessed by particular thinkers. But precisely because these empirical conditions can vary across thinkers that share the same concept, these empirical conditions are not what constitute the nature of the concept itself. A thinker's ignorance of these empirical matters of mental representation does not impugn her knowledge of which concept is in question when she thinks 'I judge that New York is hot in the summer'.

It is a striking feature of the canonical concept of a concept that it has two characteristics whose coinstantiation rests on a merely empirical truth. The canonical concept has the individuating properties specified in the leverage account. Our minds and conscious states are also such that we can rationally

apply the canonical concept of a concept in response to conscious states, such as action awareness and passive thinking, whose content involves the very concept of which it is a canonical concept. It seems to be a precondition of rational, critical thought that these two characteristics go together. Rationality requires us, on occasion, to consider for instance whether our conscious judgement that Fa was made in an epistemically responsible fashion. Investigation of this issue involves drawing on our tacit knowledge (or tacit partial knowledge) of the condition for something to be F. Such a rational exercise of thought is possible only because the canonical concept of F is one we can apply in rational response to conscious mental states whose intentional content contains the concept F itself.

Canonical concepts of concepts are far from the only concepts some of whose distinctive applications rest on empirical facts. The way we think of a type of bodily movement, when we perceive it made by someone else, yet also perceive it as an action of a type that we ourselves could make, provides another type of example. No doubt the underlying ground of the possibility of such concepts involves the now-famous 'mirror neurons' identified by Rizzolati and his colleagues. It is an empirical matter that there are such representations in our psychology. They make possible much that would not otherwise be possible. To deny the existence of ways of coming to apply concepts that rely on empirical facts would rule out large tracts of human thought and experience. This applies equally to our ability to know about the intentional content of our own mental actions and our other conscious states.[15]

[15] I thank Lucy O'Brien, Matthew Soteriou, and two anonymous referees for OUP for helpful comments.

11

Mental Actions and the No-Content Problem

Lucy O'Brien

1. INTRODUCTION

In this chapter I want to suggest that a consideration of mental action reveals a certain kind of externalism about mental content to be more problematic than we have come to think. The externalism I have in mind is generally called object-dependent externalism. The object-dependent externalist about demonstrative thoughts (ODE) I have in mind holds three theses:

1. The Object-Dependence thesis: there are demonstrative thought contents such as 'That glass is heavy' which are object-dependent. If there is no object demonstrated, then there is no content.
2. The Unity of Content thesis: there is just one content that characterizes our thought. A thought, or a constituent of a thought, cannot, for example, be constituted out of a pair of parallel contents.[1]
3. Dependence of Thought on Content thesis: all thoughts are constituted by contents. If there is no content there is no thinking.

It follows from these three theses that there can be object-dependent demonstrative thoughts: thoughts that are such that had there been no object of the relevant kind, in the relevant relation to the subject, there could have been no thought. So, if were no glass to be the object of the demonstrative thought 'That glass is heavy' there would be no content for the thought to be constituted from, and so no thought either.

If the argument of this chapter is right this version of externalism, in contrast to kind-dependent externalism, is a hard view to maintain in the face of a serious consideration of what is going when there is content failure in the case of mental action.

[1] The unity thesis is not intended to rule out the possibility that different parts of a thought might be characterized by contents classified in some way as different types. It is rather intended to rule out the dual content theorist who holds that dual and fundamentally distinct contents could be responsible for the same parts of a thought. See McGinn 1982.

2. THE NO-CONTENT PROBLEM

Let us take a relatively unproblematic example of a mental action:

(1) A supposing that *P* (where A is a subject and *P* a content).

Those who believe that there are no mental actions will obviously not accept that a supposition is a mental action. However, for most who accept that there are at least some mental actions, suppositions will seem an unproblematic example of one. Unlike judgements, or denials, suppositions are often voluntary. I can suppose that *P*, in order to work out, for example, what I would do in circumstances in which *P* was true, simply by willing to suppose that *P*.

However, in order for me simply to suppose that *P*, certain conditions have to be in place. In particular, whatever conditions are required for there to be a content *P* have to be in place. It is by now a familiar thought, indeed almost an orthodox one, that there are external conditions on content that can fail. Let us suppose then that *P* is a content with external conditions that can fail. There are two sorts of content that have been thought to be dependent upon external conditions: kind-dependent contents (such as, 'Water cleans') or object-dependent contents (such as 'That glass is heavy').

In this chapter I will consider as an example the following putative case of an object-dependent demonstrative thought: A supposes that *that* glass is heavy. We are familiar with action failure in the case of physical actions. It can seem to a subject that she raised her arm, when in fact she failed to. The suggestion often made is that if we take tryings to be antecedents of actions we can explain such failure by appealing to the idea that while the subject failed to raise her arm, she nevertheless tried to raise it. And her trying to raise it is what explains why she thinks that she raised it when she did not. (Note that there is similar story in terms of prior intentions or motor intentions: the subject knows what she intended to do and the explanation of it seeming to her that she raised her arm is to be explained in terms of her intention to move her arm.)

Now consider a case of action failure in the case of supposition. Suppose that at t, A makes the object-dependent supposition that *that* glass is heavy, and that at a later time t', it seems to A just as it seemed to her when she made the object-dependent supposition. Suppose, however, that at t' there is no object available bearing the required relation to A, and thus that there is no content for A to be supposing true. It just seems to A that she is supposing that *P*, but she can be doing no such thing: there is no content *P* available in the situation for her supposing to be a supposing of.

Faced with a failed action of this kind it is clear that we cannot make the move that was suggested in the case of failed physical action. We cannot explain the action by saying that A carried out a prior action of trying to suppose that *P*.

If there is no possibility of supposing that *P*, due to content failure, then there is no possibility of trying to suppose that *P*—due to content failure. For one needs to be able think the content *P* if one is to try to bring it about that one supposes *P*, just as much as one does if one is to suppose *P*. So, it is a striking thing about failures of content in the case of mental action that there does not seem to be the possibility of retreating to a trying or intention to explain the failure.

3. ACTION FAILURE DUE TO CONTENT FAILURE

So, the question now arises: if not in terms of tryings or intentions, how *should* we explain what is happening when it seems to a subject just as it does when she has supposed that *P*, but has not due to a failure of content?

The problem as it arises for mental actions, with object-dependent mental contents, is a version of the broader problem as it arises for any *occurrent* mental phenomena with object-dependent content. The postulation of object-dependent content allows for there to be failures of self-knowledge. It allows, as we might put it following the perception literature, for there to be good cases and bad cases.

(i) In the good case: it seems to A that she in supposing that *P*, and she is supposing that *P*.

(ii) In the bad case: it seems to A as if she is supposing that *P*, but she is not supposing that *P*. In fact there is no content indicated by '*P*'.[2]

There are two kinds of question one can ask when presented with the bad case. One can ask *epistemic questions* of the kind: How does A know that she is supposing *P*, in the good case, if she cannot rule out being in the bad case? Does A have an epistemic justification for whatever thought she does have in the bad case? Such epistemic questions have been the focus of much of the literature on externalism.

But one can also ask o*ntological questions* of the kind: What is going on in A's mental life when this kind of failure of self-knowledge takes place? An act of supposing that *P* is, I have assumed, a dated event the initiation over which I have control: I can decide to suppose that *P* and go ahead, right

[2] Perhaps the object-dependent theorist might deny that it can even *seem* to A that she is supposing that *P*, if there is no content *P*. So they might complain that they are being asked to explain an appearance that their account does not allow for. However, the object-dependent theorist must have some way of characterizing how it is in the bad case. And the bad case just is a case of failed action in which it seems to the subject as if they did what they would have done, had they succeeded. So, I will take the locution 'it seems to A that she is supposing that *P*' as not implying that A need to be able to think *P*. To do otherwise would, it seems to me, leave us no way of conducting the discussion.

then and suppose that *P*. But given that in the bad case there is no content *P*, the acts of supposing *P*, deciding to *P*, and trying to suppose *P* cannot take place—but what psychological event does take place in cases of content failure? And if no psychological event takes place, how are we to explain what is going on in the bad case? It is these ontological questions that I am interested in here.[3]

3.1. First Attempt: Thinking one is Supposing that *P*

Perhaps we can explain what is going on by appealing to a higher order activity, say the activity of thinking that one is supposing:

(A) A can think that she is supposing that *P* when in fact she is not supposing that *P* (because there is no content *P*).

But, this is no better than the appeal to trying to suppose that *P*, that we rejected above. If there is no content *P*, A cannot suppose *P* and A cannot think that she supposes *P*. Therefore, if there is no content expressed by '*P*', then (A) cannot be the right way to capture the failure: it is false or meaningless. As Bell has put it:

> The difficulty crudely speaking, is that either the non-existence of the embedded, merely apparent thought will contaminate the second-order thought of which it is a part, or, conversely, the intelligibility of the second order thought will bestow respectability on its first order component. (Bell 1988: 51)

3.2. Second Attempt: A Gap?

Perhaps the assumption that there is an act of mind—a psychological event—that constitutes the failure is an unwarranted one? There is a familiar distinction between failures of *omission* and failures of *commission*. We may be making the mistake of assuming that we have here a failure of commission, rather than one of omission. We seem to have assumed that there must be something that A does which constitutes her failure to suppose the *P*, rather than simply that A fails to suppose *P* and that therein lies her failure. Perhaps in cases of content failure there is a gap in the psychological stream where otherwise an act of mind would have been. Perhaps McDowell has something like this in mind when he says:

> there can only be a gap—an absence—at, so to speak, the relevant place in the mind—the place where, given that the sort of de re sense in question appears to be instantiated, there appears to be a specific de re sense (McDowell 1984: 103; quoted in Bell 1988)[4]

[3] David Bell's way of approaching the issue (in Bell 1988) is distinctive in asking these ontological questions, rather than the more familiar epistemological ones.

[4] McDowell himself may only have been concerned to point to a gap in the senses, and not a gap in the acts of mind.

The problem is that we have in such cases not only a failure to suppose *P*, but we have a subject to whom it is *as if* she supposed *P*, and we have to explain that appearance. When the subject introspects she does not notice a gap in her mental life, even while reflecting closely on how things are for her.

Further it is attractive to think that the phenomenology of our mental actions is sensitive to their contents. It is different for me to suppose *P*, than it is to suppose *Q*, and that difference seems most naturally explained by the contents *P* and *Q* and the differences between them. If it were to A as if she were supposing that *P*, even where there is no content *P*, then we would have to drop this natural story about the phenomenology of thought—at least in bad cases.

Most problematic in my view, the gap view leaves nothing to play the normative role that a contentful act can play when it is there. Even when A fails to suppose that *P*, due to content failure, it seems to her that she supposes that *P*, and she can act and infer as she would, were she were supposing that *P*. On the gap view we have no explanation of why it seems to A that she is supposing that *P*, or of her actions consequent on its seeming to her she is supposing 'that glass is heavy', or of her inferring that 'if my supposition is true, there is at least one heavy glass'. We seem to need *something* to do the normative and epistemic work—some associated act, or some remnant or degraded version of the act one gets in the good case.

It might be helpful to compare this case to the physical action case. When it wrongly seems to me as if I raised my left arm, even though I have in fact lost my left arm or lost motor control of it, we may not want to say that what we have here is a failure of commission—I wrongly thought I raised my left arm, but I did something else instead. We rather want to say that there was simply no physical action here: I thought I acted but failed to do so. We do however want to ask what occurrence in the psychological life of the agent accounts for it seeming to the agent that she moved her arm when she in fact did no such thing, and indeed could have done no such thing given that there was no arm to raise. And we would expect to be able to able to appeal some psychological event—perhaps an event that precedes or occurs concurrently with the action even when it is successful, or perhaps an event that occurs only when the action is not successful—in order to explain why the subject takes herself to have acted when she has not.

3.3. Third Attempt: Disjoint Contents

It may be suggested that there are different kinds of content involved in the good and bad case. Perhaps there are seemings to suppose in both cases, but that in the good case the seeming to suppose involves A thinking that content *P*, but that in the bad case the seeming to think involves A thinking some other content.

3.3.1. Other thoughts with other contents

There are various alternative contents one might appeal to in order to try to explain the bad case. Perhaps it seems to A as it does because of:

(i) associated existential thoughts she has, and can have even when there is content failure—perhaps the relevant thought contents are 'There is an object near me';

(ii) distinct singular thoughts she has, and can have even when there is content failure: 'That spatio-temporal region is F', say;

(iii) memory images she has of earlier perceptual experiences;

(iv) words that A visualizes (or 'audioizes') in the imagination. Perhaps A imagines seeing or hearing the contentless sentence 'That is F'.

All these suggested substituted contents meet something like the same problems. The first and most obvious of which is that it does not seem immediately plausible that alternative contents that are supposed to appear in the bad case will be phenomenologically indistinguishable from our supposing that *P* in the good case. Further, it is not clear how content failure is supposed to occasion such alternative thought contents or indeed why the occasioning of such contents is supposed to count as content failure rather than simply thinking different thoughts. Perhaps the idea is that such contents are not occasioned only in the bad case but are rather always involved in the supposition that *P*. The thought might be that it is only in the absence of the supposition *P* that the alternative thought contents play a role in making it seem to A as if she is supposing *P*. But if this is what is intended it is not clear that we are being offered no real alternative to the gap view: in the bad case things are just as they are in the good case, only with the supposition missing.

It may be said, however, that this view, unlike the gap view, provides contents that can play causal, normative, and epistemic roles in the bad case. But then a final concern with the suggestion is that we will expect these contents, which are distinct from the supposition, to have their own independent causal, normative, and epistemic consequences whether or not they are accompanied by a supposition that *P*. If they are around in the good case we need to know why their presence does not bring about a degree of disruption in that case.

3.3.2. Intentional *contents distinguished from* real *contents*

Rather than appealing to other familiar thoughts with distinct contents in appealing to disjoint contents in the good and bad cases, perhaps we can appeal to the idea that there are different *kinds* of content in the good and bad cases. We might argue for a bifurcation between *types* of contents: between merely *intentional* or *seeming* contents and *real* thought contents. Intentional or seeming contents are understood as not being object-dependent and as being distinct from, but in someway isomorphic to, real contents.

However this suggestion has three troubling features. First, it requires postulating contents of some new and unfamiliar sort. Second, it requires the postulated contents to occur only in the bad cases. But it is hard to exclude them from the good cases also. If its seeming to the subject that she is supposing that 'that is a glass of water' is explained by the occurrence of a merely intentional content in the bad case how do we know such content does not also occur in, and perhaps account for the phenomenology of, the good case? Third, the postulation of a distinct type of content offends against the unity of content thesis that I have assumed is part of the object-dependent externalist position.

3.3 3. Reflective Disjunctivism

Perhaps the most sophisticated attempt to appeal to disjoint contents in order to solve something like our problem comes from the theory of perception. According to what has been called Reflective Disjunctivism we can characterize what is going on in bad cases without appealing to distinct kinds of intentional content, but rather with reference to what is going on in the good case, and to the nature of a epistemic relation between the good and bad cases.[5] The reflective disjunctivist claims:

1. What is going on in the bad case is characterized by reference to what is going on in the good case.
2. In particular, what is going on in the bad case is identified as that which is *reflectively indistinguishable* by A from the good case.
3. A bad case is *reflectively indistinguishable* from a good case in virtue of the subject not being able to know by introspective reflection that they are not in the good case. For all the subject knows from introspection, she can judge that she is in a good case.

A more minimal characterization of the position would leave out the ability to judge claim—the claim that 'for all the subject knows, she can judge that she is in a good case'—and just have the inability claim—the subject cannot know by introspection that she is not in the good case. I will come back to the more minimal characterization, but will for now run with the characterization above as I think it has more mileage in the theory of perception. But I want to suggest that it does not have the same resources in the case of failure of content in thought.

In the perceptual case we can make the following claim:

(P) A subject who *hallucinates P* cannot know by introspective reflection that she is not *seeing P*. For all she knows from introspective reflection, she can judge 'I am seeing *P*.'

[5] See Martin 2002, 2004. See Sturgeon 2006 for a very helpful discussion of the position.

In the case of thought the parallel claim would be:

(T) A subject who *seems to think* that *P* cannot know by introspective reflection that she is not *thinking* that *P*. For all she knows she can judge 'I am thinking *P*'.

Given (P) we can make sense of how it is for the subject when it seems to her that she is seeing *P* by appeal to the fact that she could, for all she knows to the contrary, rationally judge that she is seeing *P*. The problem for (T) is that the subject's introspective reflective judgements would *also be thoughts*. So if there is no thought content in the first-order case there will be no possibility of second-order thought content required to make sense of the introspective judgement. In the hallucination case a reflective subject is able to make sense of what she is undergoing by the possibility of an ascent to a level of reflective thoughts for which the subject has content. A contentful ascent is allowed for because the content that constitutes the subject's reflective thoughts can be held to be distinct in kind from the content that constitutes her perceptions, so a lack of content at one level does not immediately imply a lack of content at the other. However, there can be no such ascent in the thought case. The following may be taken to represent the nature of ascent required in the thought case:

1. A is in a bad case undergoing a seeming thought episode E.
2. Bad case E is reflectively indistinguishable from a good case E'.

So,

3. A being in bad case E means that for all A knows on the basis of introspective reflection 'This thought episode E is a thought that *P*'.

But our problem is that A cannot think *that* thought. In fact there are possibly two reasons why A cannot think that thought. First, A cannot think '*P*' so could not judge 'This thought is a thought that *P*'. It is therefore difficult to see how to cash out the claim that for all A knows the thought episode could be a thought episode of thinking that *P*. Second, if the object-dependence externalist is right then it is not clear that A is in a position to think anything about 'this thought', never mind that it is or is not a thought that *P*. For if ODE is right there is no thought there to be demonstratively referred to. And if perceptual demonstrative thoughts about external objects are object-dependent, do we not also have a reason to suppose that introspective demonstrative thoughts about our thoughts are dependent on their existence?

Let us suppose that demonstrative thoughts about thoughts are object-dependent, and that judgements of indistinguishability with respect to thoughts involve reference to demonstrative thoughts about thoughts. Now, since a subject lacks a thought when the thought lacks a content, a subject can make a judgement of indistinguishability with respect to the thoughts only when she *has* real and not seeming thoughts. Therefore, only in the good case is the subject able to

make the relevant indistinguishability claims. The subject seeming to make such judgements in the bad case will in fact be having no thought at all. It may seem to her that she is thinking, and she may seem to judge that her thought is indistinguishable from the last time she thought *P*. But she is neither thinking *P*, nor thinking that her thought is indistinguishable from a previous thinking that *P*.

Perhaps we can meet the problem by embracing disjunctivism not just about the contents of first-order thoughts, but about the indiscriminabilty judgements also? Appealing again to the distinction between good and bad cases we have:

> The good case: If A thinks that T and thinks that T', A can judge that T and T' are indistinguishable. (Say, T is the thought 'That1 glass is heavy' about glass 1, and T' is the thought 'That2 glass is heavy' about glass 2.)
>
> Bad case: Of A's seeming thoughts T and T', A can seemingly judge that T and T' are indistinguishable. A's seeming judgement that T and T' are indistinguishable, is reflectively indistinguishable from A's judgement that T and T' are indistinguishable . . .

But of course, there can be no such judgement. If the problem is a real one in the ascent from the first to second order it will be a real one in the ascent to indiscriminabilty judgements also.

In raising these problems we have assumed a version of reflective disjunctivism on which it is a problem that the subject could not make a second-order indistinguishability judgement with reference to which we can explain how things seem to her. But, as was suggested earlier, there is room for a more minimal characterization of the reflective disjunctivist's position.

The reflective disjunctivist may deny that they need to appeal to the subject's higher order judgements in the bad case. All that is needed is that that the subject cannot know by introspection that they are not supposing *P*. And the subject who cannot think *P* meets that negative condition because they cannot think that they suppose *P*, and so cannot know that they are not supposing *P*.

However, it is not clear, if that is *all* that is said, that we been provided with any way of distinguishing between the case where there is a failed supposition and the case where there is no act of the mind at all. In both cases there is nothing to know and nothing known. But how do we escape the intuition that when we fail to make a supposition that *P*, due to content failure, there is nevertheless some dateable occurrence in the mind that is distinct from the case in which one does nothing at all. The kind of higher order thought we have found problematic in the bad case would enable us to make the distinction.

However, while the reflective disjunctivist cannot appeal to reflective judgements in the bad case they can appeal to those a subject might make from within the good case. Perhaps we can explain how it seems to the subject when she seems to think that *P*, but doesn't, by saying that her mental occurrence has a character which is such that she does not know she is not in a good case, and

which were she in a good case she would be able to judge her current mental occurrence as indistinguishable from the mental occurrence in the good case.

Let me say two things about this suggestion. (1) It does not get around the thought that we need a characterization of the bad case that a reflective, sufficiently conceptually sophisticated subject *can* appeal to *in the bad case*, even if she does not need to. In the perceptual case the subject suffering from a hallucination who wonders whether she is so suffering may be able to give some characterization of the state she is in by referring to the contents of her experiences when she has a veridical perception. However, allowing her that requires that she has concepts and modes of reference that she can use in both cases. We cannot meet that requirement in the case of thought.

There is a move that the reflective disjunctivist might make in response to this. Perhaps they will say that while, in the bad case, the subject is not in a position to think the particular content *P*, she is in the position to grasp the idea of a demonstrative thought of a perceptually presented glass and that the indistinguishability claims can be made in reference to such a thought. However, this suggests that the reflective disjunctivist will need to acknowledge that we need to type demonstrative contents in a way that abstracts from the particular object referred to. This acknowledgement is in tension with ODE.[6]

(2) There is an important difference in the resources of the reflective disjunctivist about perception and the reflective disjunctivist about thought. The reflective disjunctivist about perception can allow reflective indistinguishability judgements from the perspective of the bad case. For all the subject knows when suffering from a hallucination she could in those circumstances judge that she is seeing *P*. The case allows a conceptually sophisticated and reflective subject who wonders when she is hallucinating, whether she is hallucinating, to characterize how things seem to her in terms of how they seem to her when she is seeing. The reflective disjunctivist in the object-dependent thought case cannot make such an allowance. We have thus found a reason to prefer the first version of reflective disjunctivism, and found no reason, as reflective disjunctivists about perception, not to.

3.4. Fourth Attempt: Appeal to Sentences

3.4.1. Metalinguistic appeal

David Bell, in considering the problem about how to account for how things seem to the subject in the case of failures of content in object-dependent thoughts, says the following:

> How can I express the way things seem to me? Perhaps the best option is something like:
>
> (5) It seems to me that I am thinking a thought of a kind expressible by the sentence 'a is F'.

[6] Thanks for Matt Soteriou for making this point.

Although it has the virtue of concerning an apparent thought that is not a real thought while itself remaining a meaningful sentence, it nevertheless suffers from one basic shortcoming. It is false. (Bell 1988: 52)

It is false, thinks Bell, because when there is content failure it does not seem to the subject as if she is thinking about a sentence: the phenomenology does not have a metalinguistic character. It is hard not to agree with Bell on this. When it seems to me that I am thinking 'That glass is heavy' and I am indeed thinking that glass is heavy, it is highly implausible to suppose that capturing how it seems to me involves me referring to a sentence. And if there is no such reference in the good case, we would hardly fail to notice if it were a critical element in characterizing the phenomenology of the bad case.

3.4.2. Rehearsing sentences

Subjects can stand in a number of different relations to sentences. A subject can hear or see a sentence; further a subject can retain an image of hearing or seeing a sentence. And as was just brought out, a subject can refer to a sentence. These relations have not proved particularly useful in dealing with how we might characterize the phenomenology in cases of content failure. (See 3.3.1 and 3.4.1 above.) But there is another more direct relationship in which a subject can stand to a sentence. A subject can utter a sentence. And correlative to uttering a sentence by speaking it a subject can rehearse a sentence in thought. One can silently run through or speak a sentence to oneself. And as Pryor puts it: 'Rehearsing sentences to yourself is one way of having occurrent thoughts' (2006: 329 n. 1).

When the sentence is meaningful one runs through the sentence and thereby thinks a thought. But perhaps rehearsing a sentence (or syntactic string) to oneself is also one way of *not* having a thought. When the sentence does not express a thought or is meaningless one may run through the sentence but not thereby think anything—or at least not think the thought expressed by the sentence. The suggestion, then, is that we can deal with problems raised by cases of content failure by claiming that what we have in such cases is a subject rehearsing a meaningless sentence in thought.[7]

The suggestion has a couple of distinct advantages. First, it can draw on the resources of a familiar way of explaining how someone can act in a way that they

[7] Adrian Haddock suggested to me that the disjunctivist may be able to make use of the metalinguistic move in the following way. As a variant of (T) we might have:

(T') A subject who seems to think that *P* cannot know by introspective reflection that she is not thinking that *P*. For all she knows from introspective reflection the subject is able to judge that she is thinking a content expressible by 'I am thinking *P*'.

This is an interesting suggestion. It is a variant on (T) that does not leave an opening for Bell's worry about getting the phenomenology wrong as it only claims that her seeming to think it is constituted by this fact about her epistemic position. However, my suspicion is that this version of disjunctivism will suffer from the same problem we identified earlier: it does not supply the act of mind.

take to be meaningful, but which is in fact not. When someone is just starting to speak a language it is highly probable is that they will produce what is in fact a meaningless string in attempting to say something meaningful. It is arguable that small children do this for months. And it is easy to identify what event occurs in place of the meaningful speech act the agent thinks they have carried out. Suppose I take myself to have said 'It is a pleasure to meet you' in French, but have in fact not. Suppose I have in fact said nothing; I have just produced a burble. It remains the case that I have done something—I have produced a meaningless string of noises, which I take to be the speech act. If I can think by silently rehearsing sentences or quasi-sentences then we may have the beginnings of an explanation of what event occurs when someone takes themselves to be uttering a sentence but is not. When someone takes themselves to be thinking 'That glass is heavy' but is not, what they are doing is something like silently rehearsing the string 'That glass is heavy' which is meaningless.

We have seen that this suggestion has two distinct advantages. First, it provides us with an event: the rehearsing of a syntactic string that can exist even when there is content failure. Second, it is plausible to claim that the phenomenology of uttering a meaningless sentence could be the same as uttering a meaningful one. Indeed given that we are concerned—in the case at hand—with the very same syntactic string being meaningful or meaningless depending on external conditions we have a reason to think that the phenomenology will be the same.

I have, however, four problems with the suggestion. First, in the case where we have a speaker uttering a meaningless string, and thus failing to say anything by the action produced, we are inclined to say that in the case of failure there was nevertheless something that the speaker was trying to say. In the example, above, I was trying to say 'It is a pleasure to meet you' in French when I produced a meaningless burble. If an account of what the subject is trying to do is going to presuppose that the subject be able to think a thought with a content that in some way matches or correlates to what she is trying to do—as seems natural—then appealing to this model in the failed thought case is going to be problematic. I can try to say 'It is a pleasure to meet you' in French because I can think 'It is a pleasure to meet you' in English. However, in a case of content failure, I cannot *try* to think '*That* glass is heavy' in rehearsing a syntactic string, in virtue of thinking '*That* glass is heavy'. If I cannot think the thought, but only try to, that cannot be in virtue of any capacity to think it. In the case of failed physical actions we have our thoughts about what we are up to to fall back on; in the case of acts of thinking we do not.

It may be that the model we have appealed to in the failed speech case is not the right one—perhaps trying to say *P* does not in fact presuppose being able to think *P*. If that is the case, we will be lacking an account in both the failed speech and failed thought case. However, it is surely natural to think that trying to F implies the capacity to have a thought involving F, and natural to think that there is a disanalogy between the two cases.

The second problem with the suggestion that we explain failed thought on the model of failed speech is that while it seems right to say, as Pryor does, that rehearsing a sentence is one way of thinking a thought, it is highly contentious to think it is the *only* way. At least the phenomenological evidence does not prima facie support the general thesis: we do not *seem* to be silently talking to ourselves whenever we are thinking. Given that, it is perhaps implausible to take this suggestion as a general explanation of failures of content.

The third problem is that the suggestion does not seem to provide a fine-grained enough solution to the problem of content failure. Consider the case of a subject perceiving two similar glasses: one to her left and one to her right. Suppose she forms a demonstrative thought about each glass. '*That1* glass is heavy' and '*That2* glass is heavy'. She does so, one presumes, by attending to, mentally nodding towards each glass. These are distinct thoughts, so if thinking them involves something like the rehearsing the sentences used in their expression then something other than the sentence rehearsed is required to individuate the thought: the sentences themselves are not syntactically distinct. Also required for distinguishing the thoughts is the picking out, focusing on, attending to, the objects in order for the thought to be fully completed.

But now consider a complex hallucination that is phenomenologically indistinguishable from the case just considered. In this case there are two instances of content failure—the subject both fails to think 'That1 glass is heavy' and fails to think 'That2 glass is heavy' but it seems to her just as it seems to the subject in the good case. That is, it seems to her that she is thinking two thoughts and it seems to her that they are distinct thoughts.

The appeal to silently rehearsing syntactic/sentence types is not sufficient to explain what is going on in such a case. Given that the sentences rehearsed are identical in the two instances, they are going to underdetermine the phenomenology. One possibility would be to hold that, over and above the silent rehearsal of a sentence, what we have in these cases are distinct acts of attending: on the one hand to what seems to be a glass on the left, and on the other to what seems to be a glass on the right. However, if we are not going to reduplicate our problem as applied to such acts of attending then we are going to have to allow for attentional content, or perceptual content that is not object-dependent. Thus the ODE about thought cannot also be an ODE about attentional content or perceptual content. However, it is very difficult to see what motivation there could be for insisting that there is no content that would characterize the thought if there *were* a content that characterized the act of attention or perception.

The fourth and final problem for the suggestion that in cases of failed thought what we have are silent rehearsals of a sentence, concerns how the suggestion fits with the broader profile of ODE. In particular, there is a question of whether, given certain plausible assumptions, it is consistent with the Unity of Content Thesis and the Dependence of Thought on Content Thesis that were identified as commitments of ODE.

Let us take the Dependence of Thought on Content thesis first. The claim is that we fail to think a thought, but rather silently run through a sentence. But what is silently rehearsing a sentence if not a kind of thinking? And if silently rehearsing a sentence is a kind of thinking then there is on the suggestion contentless thought. I think ODE can deal with this quite easily. Consider the case of speech. It is plausible to say that I said nothing when I tried to say 'It is a pleasure to meet you' in French, but produced a bit of burble. However, it is also very natural to say that my burbling was a kind of speaking—a deliberate production of my organs of speech. Given the possibility of a similar such ambiguity in the case of thought, ODE can very reasonably suggest that the thinking involved in rehearsing a contentless syntactic string is not the kind of thinking at issue in the Dependence of Thought on Content Thesis.

What of the Unity of Content thesis? The issue here is tricky and a proper resolution of it will not be attempted. However, there is an obvious worry that in taking up the 'inner utterance' suggestion the ODE will be taking a step closer to accepting a dual content thesis for thought. Consider again the case of a subject perceiving two similar glasses, one to her left and one to her right, and forming a demonstrative thought about each glass. '*That1* glass is heavy' and '*That2* glass is heavy'. Suppose that the ODE appeals to the idea of the subject rehearsing a meaningless sentence, to explain what is going on in the case in which the subject is having a complex hallucination that is phenomenologically indistinguishable from the case just considered. Then what reason does ODE have not to take the subject to be rehearsing a sentence in the good case also? And if the subject *is* rehearsing a sentence in the good case also it is extremely plausible that, like uses of the demonstrative 'that' in English, what we have is a general content that determines a context-dependent content in use. If this is right, then while there may be distinct, perspectival singular contents characterizing each thought, such contents will be dependent upon the application of a general content that is grasped in understanding the general meaning of the words 'That glass is heavy'.

If there is a general content that is grasped in understanding the words 'That glass is heavy' then it appears that we have more than one kind of content to appeal to. Thus, in cases of reference failure, while there may be no distinct singular contents characterizing each thought, there will be distinct acts of applying a general content that can be appealed to in the explanation of what act of mind takes place when the subject tries to think '*That1* glass is heavy' and '*That2* glass is heavy'. However, this explanation will be at the cost of the Unity of Content thesis.

I will conclude by remarking that the problem that has been explored here for the Object-Dependent Externalist does not seem to be as serious a problem for an externalist who is only a Kind-Dependent Externalist (KDE). Consider a KDE who holds that there are natural-kind thought contents such as 'Water is in lakes' which are kind-dependent. If there is no kind, there is no content. Suppose a subject is on Dry Earth. She thinks she has a concept associated with

the word 'water' but she does not: there is a systematic mirage meaning that all the taps, lakes, and rivers she takes to be filled with a stuff she calls 'water' are in fact empty. According to the KDE it might seem to her that she thinks 'Water is in lakes' but she can she think no such thing: there is no thought available to be thought. Suppose, however, KDE allows that she can have demonstrative thoughts such as 'that stuff coming out of the tap' or 'that stuff in the lake' even when there is no stuff demonstrated.

What explanation can KDE give as to what happens in cases of content failure? What happens when it seems to our subject that she thinks 'Water is wet' but in fact thinks no such thing, there being no concept water to think? Consider again the suggestion that she rehearses a syntactic string 'water is wet' which lacks content. KDE's use of the suggestion is not problematic in the same way that ODE's use is. In the demonstrative case this explanation was not sufficient because of the possibility of distinct failed thoughts having the same syntactic structure. In the non-demonstrative case there is no such impediment to the syntactic structure being sufficient to determine both the thought, and the phenomenology of the failed thought. Further, there is no reason to take the Unity of Content thesis for kind thoughts to be in tension with the 'inner speech' suggestion. If our subject means nothing by the thought word 'water' in this context she will mean nothing by the thought word 'water' in another she is taken to. Given that there are no contexts in which her repertoire of words and concepts enables her to use 'water' contentfully (in contrast to her use of the demonstrative 'this stuff') we have no reason ascribe to her any understanding of a general content.

I conclude that when we focus on what is happening in the mind of a subject in an instance of content failure we find that the object-dependent externalist is pushed to explain what is going on. The kind-dependent seems to be in a better position.[8]

REFERENCES

Bell, D. (1988) 'Phenomenology and Egocentric Thought', *Proceedings of the Aristotelian Society, Supplementary Volume,* 62: 45–60.

Evans, G. (1982) *The Varieties of Reference* (Oxford: Clarendon Press).

McDowell, J. (1984) 'De Re Senses', in C. Wright (ed.), *Frege: Tradition and Influence* (Oxford: Blackwell).

[8] An unwritten and much earlier version of this chapter was presented to the 'After Kant' conference in Sheffield, marking the retirement of David Bell. Many thanks to David, and the organizers, for occasioning me to go back to Bell 1988. Thanks also to the audience in Sheffield, and to one in York, for very helpful comments and questions. Thanks for very helpful comments from Adrian Haddock. Discussion with Matt Soteriou on a number of occasions, from the start, has been hugely helpful—indeed essential. Many thanks to him. Finally, my gratitude to Mark Sacks for having, this time and always, helped me work out what I was thinking by talking with him.

McGinn, C. (1982) 'The Structure of Content', in A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press).

Martin, M. G. F. (2002) 'The Transparency of Experience', *Mind and Language,* 17: 376–425.

——(2004) 'The Limits of Self-Awareness', *Philosophical Studies*, 120: 37–89.

Pryor, J. (2006) 'Hyper-reliability and Apriority', *Proceeding of the Aristotelian Society,* 106: 329–46.

Sturgeon, S. (2006) 'Reflective Disjunctivism', *Proceedings of the Aristotelian Society, Supplementary Volume*, 80: 185–216.

12

Mental Agency, Conscious Thinking, and Phenomenal Character

Matthew Soteriou

The ordinary, common-sense notion of action is one we most naturally associate with overt bodily actions, as opposed to mental actions like calculating, trying to remember and certain kinds of imagining, and one finds this reflected in the emphasis placed on bodily action in the philosophical literature. Discussions of action usually only mention mental action in passing, if at all, and, more generally, the topic of mental action has received relatively little attention in work in the philosophy of mind. A notable exception is Brian O'Shaughnessy's insightful treatment of the topic in his book, *Consciousness and the World.* Through his various discussions of mental action and 'mental will' in this book, O'Shaughnessy makes a case for thinking that an exploration of issues germane to mental action in particular has the potential not only to illuminate and inform our accounts of action in general, but also to illuminate certain debates in the philosophy of mind that don't usually fall under the heading 'the philosophy of action'.

One of the central themes of the book is that acknowledging the extent to which one's conscious mental life is *agential* in character is crucial to an adequate understanding of consciousness. O'Shaughnessy writes, 'In the final analysis it is because thinking is active and thinking is essential to consciousness that mental action is a necessary condition of consciousness' (2000: 264) By 'consciousness' O'Shaughnessy means wakeful consciousness, and in the book he focuses in particular on wakeful consciousness in self-conscious beings.[1] O'Shaughnessy's approach to consciousness is somewhat unusual. Much of the

[1] O'Shaughnessy (2000: 89) writes, 'The mind of one who is conscious is necessarily a mind actively governing the movement of its own attention and thinking processes . . . In general the direction taken by our thoughts and attention is in the conscious actively self-determined . . . the conscious find themselves in the grip of a *necessity* to freely choose their own occupation of thought and attention. You might even say that we awake in this world in mid-stream, swimming for our lives', and 'With the imposition of a "will freeze" upon conscious experience, one cannot but replace the prevailing state of consciousness, waking, with another state of consciousness, perhaps sleep' (p. 229).

contemporary literature on consciousness is vexed with the task of understanding the nature of phenomenal consciousness—the 'hard problem' of consciousness. These discussions usually focus on the most striking examples of aspects of mind that have phenomenal properties—that is, on perceptual experiences and sensations—aspects of mind that we tend not to think of as agential, and there is very little discussion of the relevance that the topic of mental action might have to our understanding of phenomenal consciousness. In this chapter I shall follow O'Shaughnessy's lead and focus on the more neglected agential aspects of mind—aspects of mind that O'Shaughnessy claims are essential to consciousness as he construes the term—with the aim of exploring ways in which this might illuminate our understanding of phenomenal consciousness.

The outline of the chapter is as follows. I'll start by distinguishing a variety of mental action in terms of the restrictions on our third-person access to it. I'll then focus on the question of how, given this restriction, we should accommodate the agent's perspective on his or her own mental actions. Ultimately I'll argue that such mental actions must involve phenomenally conscious mental events if the agent is to have the appropriate perspective on them.

I

Specifying what it is that makes an action a mental one, as opposed to a physical or bodily one, is not a straightforward matter, particularly if one thinks that all mental events are bodily events. And, of course, any problems there may be in specifying the distinguishing mark of the mental in general will infect a debate about what it is that makes a mental action *mental*. A host of questions arise when one starts to think seriously about what it is that distinguishes mental action from bodily action—some of which may threaten to collapse the distinction.[2] For instance, if all mental events are bodily events, then aren't all so-called mental actions really bodily actions—bodily events that are intentional under some description?[3] One might also try to collapse the distinction from the other direction: When an agent performs a so-called bodily action, doesn't she *try* to perform the action? Is this attempt—this trying—to be thought of as a mental event—a mental action?[4] If so, isn't it also correct to say that the agent's attempt at performing the action is partly constituted by certain of their bodily movements, and so shouldn't this lead us to concede that all so-called bodily actions are in fact mental actions?

[2] In this chapter I do not discuss the possibility of distinguishing physical actions that do involve bodily actions from those that do not. For a discussion of this point, see the discussion of miracles in Ruben 2003.

[3] Compare Davidson's (1971) discussion of cerebral events involved in action.

[4] See e.g. O'Shaughnessy 1974, and Hornsby 1997: essays 5 and 6.

There does not seem to be an obvious and uncontroversial way of classifying every action as either mental or bodily. And it is not obvious that the categories are exclusive. Are there purely mental actions—perhaps trying to work out some arithmetical problem in your head? Are there actions that involve both bodily and mental actions—perhaps trying to figure out how to put together a self-assembly desk? Are there actions that are both mental and bodily—perhaps thinking out loud?

Philosophers sometimes sidestep such complications by simply introducing the notion of mental action by example, and some of the more frequently cited examples are calculating in your head, various kinds of imagining, deliberating, and trying to remember. What is common to these kinds of activities, one might think, is the fact that they do not seem to require for their successful performance the performance of any *overt* bodily action. We might, then, appeal to this feature of these activities, in order to distinguish at least a variety of action that we tend to think of as mental. It is this variety of mental action that I intend to focus on here.

To call a class of actions overt suggests that they are being distinguished by the kind of third-person access we can have to them. It suggests that their distinguishing feature is that they are observable.[5] Let us suppose, for the sake of argument, that all mental events are in fact bodily events. This may suggest that whether a particular action is classified as overt or covert is simply going to depend on the contingent matter of how difficult it happens to be for us to observe the bodily events that constitute it. However, in the case of the kinds of mental actions I just mentioned—calculating in your head, deliberating, various kinds of imagining, and so on—even if the mental events that constitute such actions are identical with bodily events, what is distinctive of such actions is that one can only identify such bodily events as agential by first identifying them with mental events that cannot, *as such*, be observed.

The claim here is that our third-person access to the variety of mental action I have in mind appears to be importantly distinct from our third-person access to what we usually think of as overt bodily actions. This is not to deny that one can know what another is doing when they are engaged in some such mental action. One can, for example, be told by the agent what they are thinking or imagining. The point is, rather, that there is a way of knowing what another is doing in the case of overt bodily action that is not available in the case of such mental actions. Even if one knows someone so well that one can often tell what they are thinking, it seems wrong to claim that one acquires this knowledge by observing the mental acts that constitute their thinking or imagining.

One can sometimes see *that* someone is thinking, without knowing what he or she is thinking, and equally, one can sometimes see that someone is performing

[5] Compare Dennett's speculations on the development of *sotto voce* talking to oneself as achieving a certain useful privacy (1991: 197).

some 'overt' bodily action without knowing exactly what he or she is doing. But even here there is still an important distinction to be made—a distinction that can perhaps be brought out by considering the differences between the question, 'what are you doing?' asked of the overt bodily action, and the question, 'what are you thinking?' asked of the mental action. In the former case one may be able to observe the agent's bodily behaviour as agential and wonder about the intention with which it is being performed.[6] In the latter case, one is not similarly able to observe the agent's mental events, as such, and then wonder about the intention with which they are being performed.[7] When one is given an answer to the question, 'what are you doing?' asked of the overt bodily action, then perhaps the observable bodily behaviour can start to make sense in way that it didn't before, or perhaps, given what one is able to observe, one can point out to the agent of the action that what they are doing may not be the most effective way of achieving that aim. These points do not seem to apply in the same way to the case of mental action when one is simply told of the agent's aim in engaging in that mental action.

So to summarize, I am distinguishing a variety of mental action in terms of the kind of third-person access we can have to its instances. These are mental actions whose successful performance does not require the performance of any overt bodily action. And I have said that an action is an overt bodily action if one can have the following kind of third-person access to it: one can identify the bodily events that constitute it as agential without having to identify them with mental events that cannot, as such, be observed. In the next section I consider how this restriction on our third person access to these sorts of mental actions can affect how we conceive of them.

II

If we cannot observe, as such, the mental events that (partly) constitute a mental action another person is engaged in, then it seems we should accept that either the mental events that constitute such a mental action are observable as such, but only by the agent of the action, or that the mental events that constitute such a mental action are not, as such, observable by anyone at all—including the agent

[6] On the observability of agential aspects of bodily behaviour, see Marcel 2003: 55. See also Hornsby 1997: essay 6.

[7] So the suggestion here is that from the fact that one sees that *x* is Φ-ing, or sees *x* Φ-ing, it does not follow that one thereby sees the event that is *x*'s Φ-ing (just as from the fact that one watches *x* Φ-ing, it doesn't follow that one thereby watches *x*'s Φ-ing). For example, suppose that a subject is asked to solve an arithmetical problem in his head. One may be able to see that he is attempting to solve the problem, one may be able to see him, and watch him, attempting to solve the problem, but it doesn't follow that one thereby sees the conscious mental events that constitute his attempt to solve the problem.

of the action. The former suggestion might look like an invitation to adopt the much-criticized 'Cartesian Theatre' model of mind. According to such a picture, self-awareness is to a certain extent modelled on other-awareness, but where the difference between these two forms of awareness resides in the fact that in the case of self-awareness the entities observed are private to a particular subject whose access to them is infallible. There is much in such a picture that can be, and has been, criticized. For now, I want to focus on what may be involved in taking the second option—i.e. the claim that the mental events that constitute such a mental action are not, as such, observable by anyone at all.

On this latter view one might think of the mental events that constitute a mental action as being more like theoretical entities than observable entities.[8] The idea here is that we not only attribute mental states in order to explain and rationalize overt bodily actions or behaviour, we also attribute or posit mental events that involve changes to, and transitions between, these mental states, in order to explain and rationalize the observable overt bodily behaviour that is our primary data. So according to this view, mental actions are to be thought of as a subclass of this more general class of mental processes that involve transitions between mental states.[9] The idea here is that this is supposed to be consistent with denying that the mental events that constitute a mental action are, as such, observable by anyone at all—including the agent of the mental action.

If one conceives of mental action in this way then obviously a lot more needs to be said about what it is that makes such a mental process a mental action. For surely not all transitions between mental states that enable and explain bodily action are to be thought of as properly agential. For example, the mere fact that a mental process is goal-directed is not in itself sufficient for that mental process to be regarded as a mental action, even if it is flexible and initiated by the agent's motives. For there are goal-directed sub-personal computations that we tend to think should be regarded as *sub*-agential. Such processes may play a role in enabling genuine actions, whether bodily or mental, but they are not themselves genuinely agential.[10]

[8] The allusion here is to Sellars's (1963) discussion of 'private episodes' and 'our Rylean ancestors'.

[9] In some discussions an ontological category of process is distinguished from the category of event (see Mourelatos 1978). In this chapter I do not have that distinction in mind in my use of the term 'process'. As I'm using the term here a mental process can be thought of as a series of mental events.

[10] We also tend not to think of the passive acquisition of a belief as a mental action, even if it involves some sub-personal computational procedure that somehow implements the requirements of rational belief fixation. We tend not to regard these sub-personal computational processes, whatever they may be, as *agential* processes. For a discussion of the difficulties involved in specifying such a computational procedure and the consequent significance for the computational theory of mind, see Fodor 2001. If we do think of sub-personal processes as agential, this, I suggest, is simply because of their association with, their enabling of, genuinely agential acts—whether bodily or mental. Our question is, what is it for a mental process to be one of these genuinely agential mental acts?

A natural assumption to make is that the *agent's* perspective on whatever mental processes enable and explain her bodily action has an important role to play in determining whether that particular mental process is to be regarded as properly agential. This in turn may suggest that a mental process must be, in some sense to be clarified, accessible to the agent if it is to be regarded as properly agential. So, for instance, one might think that the kind of mental process that can in principle occur during dreamless sleep is not to be thought of as agential. Such processes are not accessible to the agent in a way required for genuine mental agency. But would this be to commit to the general claim that one must always be aware of what one is doing when one is performing an action?

In the case of bodily action it is plausible to claim that we are not always fully aware of what we are doing when we are acting. We can, for example, do things absent-mindedly. But one still might regard such absent-minded actions or activities as being, in some relevant sense, accessible to the agent. In such cases one can, in principle at least, realize or discover what one is doing and subsequently attempt to control one's behaviour. So even if one thinks that the claim that an agent always knows what she is doing when performing an action is too strong, one still might think that a genuine action, whether bodily or mental, has to satisfy some kind of accessibility requirement. Compare O'Brien (focusing on bodily action) on this point:

> We can obviously fail to know that we are acting, as when we are acting absent-mindedly . . . But it does not seem to be the case that our actions can be, as a matter of brute fact, beyond our ken . . . It seems to me that we cannot, in Shoemaker's phrase, envisage a creature which is simply self-blind with respect to all their actions in this way . . .
> Surely if I have the power to initiate or stop what I am doing then what I am doing must normally be in some way accessible to me.[11]

How are we to make sense of the idea of a mental process satisfying the kind of accessibility requirement distinctive of an activity that can be properly regarded as agential—the kind of accessibility that could, for example, allow that the activity be carried out in an absent-minded way?

I've already mentioned examples of mental processes that I've suggested don't appear to satisfy the kind of accessibility to the agent that is required for genuine mental agency—i.e. sub-personal informational processing. What is missing in such cases is not simply the belief that the process is occurring. I may discover what information processing is occurring at the sub-personal level, but this does not thereby make the processes accessible in the way required for genuine agency. There's an important distinction to be made between, on the one hand, awaiting the outcome of a mental process one knows is occurring in one, and, on the other hand, being actively engaged in some mental process.

[11] O'Brien 2003: 364–6.

The kind of accessibility we are after seems to be exemplified by that aspect of our mental lives that we commonly call conscious thinking. The kind of accessibility involved in conscious thinking is not simply a matter of one's knowing that a mental process is occurring in one, and it involves the kind of accessibility that enables us to make sense of the idea of absent-minded mental action or activity—e.g. daydreaming or idle thought.[12] But what is it for a mental process to be an instance of conscious, as opposed to non-conscious thinking? The question I now want to address, in particular, is the following: if a mental process is to be an instance of conscious as opposed to non-conscious thinking, must it involve mental events or acts that have *phenomenal* properties? Sensory imagining and perceptual remembering involve mental episodes with phenomenal properties, but what of conscious calculation or deliberation? In what follows I'll be focusing on the issue of what we should say about these latter examples of conscious thinking. Note that I am not here assuming that all conscious mental episodes that are involved in conscious thinking are mental actions. Indeed I am not assuming that whenever the activity of conscious thinking occurs the agent is engaged in what should be characterized as a mental action. The claim currently being considered is just that conscious thinking satisfies the accessibility requirement that is *necessary* for mental action.

III

Conscious thinking cannot simply be a matter of the occurrence of a process that results in the acquisition of a personal-level mental state (a belief, desire, or intention, say). Such aspects of mind—personal-level mental states—can be non-occurrent. We think of them as being the kinds of features of mind that can persist during dreamless sleep.[13] And there seems to be no reason to deny that an event that is the acquisition of such a non-occurrent state can also be non-conscious and non-occurrent.[14] This suggests that there is no reason to think that we cannot make sense of the existence of a non-conscious process constituted by series of non-conscious, non-occurrent events that are simply changes to, or acquisitions of, non-conscious, non-occurrent personal-level mental states. Again, the kind of mental process that could, in principle at least, occur during

[12] Compare Walton on imagining: 'We sometimes decide on what to imagine . . . we form intentions to imagine this or that and carry them out. Imagining is sometimes deliberate. But not always. Often we find ourselves imagining certain things. Our fantasizing minds stray, seemingly at random, without conscious direction. Thoughts pop into our head unbidden . . . Varying degrees of control may be exerted over whether and what we imagine' (1990: 14). For the suggestion that we also regard dreams as actions, see Lear 1990: 71.

[13] See O'Shaughnessy 2000: 177. See also Ewing 1948: 202, and Audi 1999: 100.

[14] Compare the discussion of experiences, states, and processes in O'Shaughnessy 2000: ch. 1.

dreamless sleep.[15] Although such processes would affect an agent's personal-level mental states, that would not suffice for them being instances of conscious thinking.

This may suggest that we should go for some sort of higher order account of what it is for the mental episodes or acts that constitute one's conscious thinking to be conscious.[16] My strategy in exploring the possibility of some sort of higher order account of conscious mental acts is to raise questions about the *ontological* category of the *objects* of the hypothesized higher order states.[17] So suppose we were to opt for some kind of higher order account of the conscious mental acts/episodes involved in conscious thinking. What should we say about the ontological category of the *objects* of the hypothesized higher order states? For example, are they states or events? And if they are events are they instantaneous events, or events with duration?

If we are to accommodate the idea that an agent can apparently be *doing* something when consciously thinking, then it seems that the objects of the higher order states must include events, and not just other mental states. Are such events to be regarded as events that are the mere acquisitions of mental states with content and nothing more—instantaneous mental events that are the acquisitions of mental states with content? In Vendler's terminology, are they 'achievements'?[18] If the mental event in question is an achievement, an event that is the mere acquisition of a mental state with content and nothing more, then it seems to be the kind of event one can only have access to via one's access to the state that is acquired. For there is nothing more to the event, we are supposing, than its being the acquisition of that state.[19] So it seems that the subject of such a higher order state can only conceive of such an event as one that has happened or occurred (or as something that is going to happen), rather than something that is happening or something she is doing. And she cannot conceive of the state acquired, I have suggested, as something she is doing, as it is not an instance of the appropriate ontological kind.

[15] Compare Carruthers's criticism of first-order accounts of conscious thinking (2005: 5–6). See also Carruthers 2000: ch. 6. For a discussion of the idea that we are *always* dreaming while asleep, see Flanagan 2000: 68.

[16] As I am using the term, in committing to a 'higher order' account of conscious thinking one need not thereby commit to a higher order *representationalist* account of conscious thinking. The former simply involves the claim that higher order representations are necessary for conscious thinking and is consistent with what has been called a 'phenomenist' account of phenomenal character. See Byrne 2004.

[17] For convenience I shall continue to refer to the higher order aspects of mind as states, rather than events or states, but nothing of significance is supposed to turn on this.

[18] See Vendler 1957.

[19] Compare Mouton on 'starting': 'It is a conceptual point that starting has no duration . . . we do not consider any part of the motion or rest as being part of the starting. To move at all is to have *started*. There can by definition be no duration during which one is *starting* . . . Then the function of "start" is not to describe a process, activity or state, but to suggest that a change has taken place' (1969: 69).

This might suggest that in the case of a conscious mental act the object of the higher order state must be a mental event with duration, and not a mental event that is the mere acquisition of a state with content and nothing more. In contrast to the case of mental states with duration, we tend to think of mental events that have duration as having temporal parts.[20] Mental events that have duration unfold over time in a way that mental states do not. The suggestion so far has been that conscious thinking satisfies the accessibility requirement necessary for mental agency, and that if we go for some kind of higher order account of what it is that makes the mental acts that are involved in conscious thinking conscious, then the objects of these higher order states should be mental events with duration, aspects of mind that unfold over time in the way that mental states do not. But now we seem to reach a puzzle here—a puzzle raised in Geach's work on the mental act of judging.

It is natural to assume that the conscious mental act of judging is a prime example of the kind of conscious mental act/episode involved in the varieties of conscious thinking that are our concern—activities like conscious calculation and deliberation. According to the line of thought being currently considered, this conscious mental act should be represented by the subject's higher order mental state as a mental event with duration. However, Geach's arguments suggest that the mental act of judging lacks duration. In *Mental Acts* Geach discusses the question of the difference between speech and thought as regards temporal duration:

> Spoken words last so long in physical time . . . —one could sensibly say that the utterance of the words was simultaneous with the movement of a body . . . from one place to another. The same would go for the duration of mental images of words, or any other mental images . . .
> With a thought it is quite different. Even if we accepted the view . . . that a judgement is a complex of Ideas, we could hardly suppose that in a thought the Ideas occur successively, as the words do in a sentence; it seems reasonable to say that unless the whole complex content is grasped all together—unless the Ideas . . . are all simultaneously present—the thought or judgement just does not exist at all. (Geach 1957: 104)

Is it possible to stop a subject halfway through his act of judging? Is it possible for there to be a situation in which it was not yet true that the subject had judged that *p*, but in which it was true that he had started judging that *p*?[21] Here we might ask, what could the subject have done that counted as starting

[20] See Steward 1997: ch. 3. (See also Kenny 1963: ch. 8; Vendler 1957; Mourelatos 1978.) As Steward notes, there do seem to be cases where the obtaining of a state, which does not unfold over time, depends on a series of events, which do unfold over time. Steward gives as an example the dependence of temperature in a gas (state) on the motion of molecules (events)—p. 72. This suggests that whether we regard an aspect of mind as mental state or rather as a mental event should not simply depend on the ontological category of the physical realizer.

[21] In Vendler's (1957) terminology, what is being asked here is if the act of judging is an accomplishment.

the act of judging, and what else would he have needed to do in order to finish his judging? If we think of the content of the act of judging as having parts that signify the temporal parts of the act, then this would perhaps provide us with an answer. But Geach's observation is that we should not regard the propositional content of an aspect of mind in this way. Geach argues that the mental act of judging is individuated by its propositional content. Such contents are complex in structure, so in a sense have parts, or elements, but according to Geach the elements of the content of a mental act are not, and do not signify, temporal parts of the mental act they individuate. To borrow O'Shaughnessy's analogy, an act of judging is not like putting together the discrete parts of a jigsaw, nor is it like an artist making preliminary sketches of the final work.

The act of judging does not appear to be an activity either—i.e. an open-ended process with duration, such that it would make sense to think of a subject's act of judging that *p* as something that he had already been doing for some period of time and which he could continue to do more of.[22] There is a use of the verb 'judge' that picks out the mental *state* of belief ('S judges that *p*'), but where the verb is used to pick out a mental event, the indications are that we should regard it as picking out an achievement—an instantaneous event that lacks duration.

This leads Geach to label an act of judgement a 'non-successive unity'. Considerations of this kind lead Mouton, in his paper 'Thinking and Time', to claim that,

> The occurrence of a thought is not a durational event.
> It is impossible in principle for one to get halfway through a thought and stop. This is because thoughts are individuated by their content and hence every change in the content of one's thinking is a change of thought and every such content which comes before one's mind is a complete thought. There is, therefore, no such thing as a partial thought. (1969: 64)
> It appears to be impossible in principle for the occurring of a thought to occur in stages or to occupy any duration. (1969: 65)
> Thoughts do then occur at particular moments but we cannot in principle specify how long the occurring of a thought lasted nor how much time it took. It does not make sense to ascribe any duration to the occurring of a thought. (1969: 64)[23]

The conclusions that suggest themselves here are the following. If we assume that an instance of conscious thinking involves conscious mental acts or episodes, and if we are appealing to a higher order account to explain what it is for such mental acts or episodes to be conscious, then it seems that the object of each higher order state must be event-like—it cannot be another mental state. It also cannot be an event that is the mere acquisition of a mental state and

[22] The same considerations appear to apply to the case of the mental act of deciding to Φ.
[23] See also the discussion of Geach in Ginnane 1960.

nothing more. It seems that the object of each higher order state must be the kind of mental event that can unfold over time. Our problem is that if we take the mental act of judging that *p* as an example of the kind of mental event involved in conscious thinking, then it does not seem to be the kind of mental event that *can* unfold over time. It is not the kind of mental event that has duration.

Geach's puzzle is raised in the context of the question of the temporal relations between saying that *p* and judging that *p*. I am going to suggest that consideration of the case of thinking out loud will help us to resolve the puzzle, and then I'm going to suggest how we should apply this solution to the case of conscious thinking.

In the case of thinking out loud, say, calculating whether *p* out loud, it seems wrong to think of the out-loud utterances as overt actions that merely accompany, and that are separable from, the real mental activity of calculating whether *p*. This is because then we would not seem to have a genuine case of calculating whether *p out loud*, but rather a case of the agent saying out loud what he or she had just done.[24] In the case of thinking out loud, I suggest we should regard the overt bodily action of speaking out loud as a *vehicle* of the mental action, and not as a separable action that merely accompanies the mental action. One can be tempted into regarding thinking out loud as involving the conjunction of two separable actions—mental action *plus* bodily action—for the following reason. Saying various things out loud is not, in itself, sufficient for thinking out loud, and one doesn't need to say anything out loud in order to calculate whether *p*. So this can make it seem as though the mental action of calculating must be constituted by something other than the overt bodily action—a distinct 'inner' process, separable from it. So we have two separable actions accompanying each other—separable actions the conjunction of which constitutes a hybrid action—mental action plus bodily action. But this, I think, is a mistake. While it is true that a particular action instantiating the mental type, calculating whether *p*, could have been performed without performing an overt bodily action, and a particular action instantiating the overt bodily type, saying various things out loud, could have been performed without performing the mental action, the action also instantiates an action of a basic, non-reducible type, that we might call *mental action with an overt-bodily-action vehicle* (in this case, calculating whether *p out loud*). An action instantiating this type could not have been performed without performing both types of action—mental and (overt) bodily. According to this way of regarding the case, the verbal utterance instantiates two kinds of action—overt bodily (talking out loud) and mental (calculating whether *p*), in virtue of the fact that

[24] Compare Wittgenstein here: 'There are important accompanying phenomena of talking which are often missing when we talk without thinking, and this is a characteristic of talking without thinking. But *they* are not thinking' (1953: 218).

it instantiates a third, basic, non-reducible kind of action, namely a mental action with an overt-bodily-action vehicle (in this case, calculating whether *p* out loud). So an event of one's verbal utterance can instantiate two types of act—one's saying that *p* out loud, and one's judging that *p*, *because* it instantiates a third basic, non-reducible type of act, namely one's judging that *p* out loud.

With this discussion in mind let us return to the problem of the higher order account of a conscious mental act, such as consciously judging that *p*. The puzzle reached was the following. The object of the higher order mental state cannot be another mental state, nor can it be a mental event that is the mere acquisition of a state and nothing more. It must be an aspect of mind that has duration and also has temporal parts—something event-like that unfolds over time. But Geach's argument suggests that the mental act of judging is an achievement—an instantaneous mental event that lacks duration, so how, as O'Shaughnessy puts it, can one stretch such a thought across time?

The solution, I suggest, is to think of the conscious mental act of judging as involving the occurrence of a conscious event with duration that is the vehicle of the mental act of judging, just as, in the case of thinking out loud, the bodily action of one's saying that *p* is the vehicle of one's judging that *p* out loud. Since we are dealing here with a mental act without an overt bodily action vehicle, the apparent vehicle obviously cannot be an overt bodily action. It must be a mental action or event—e.g. one's saying something in inner speech—an event that is phenomenally conscious.[25] Although their phenomenal characters may not individuate mental actions such as calculating and deliberating, it does not follow from this that they can be carried out without the occurrence of any phenomenally conscious mental events. It is true that an event that is the acquisition of a mental state, e.g. the belief that *p*, can in principle be instantaneous and occur without the occurrence of any phenomenally conscious mental event that unfolds over time. It also true that the phenomenally conscious mental event that unfolds over time, e.g. one's saying something in inner speech, can in principle occur without the acquisition of the belief that *p*. However, an event of that phenomenal kind can, under the right circumstances, instantiate an event of *both* types (acquiring the belief that *p* and saying that *p* in inner speech), in virtue of the fact that it instantiates an event of the basic, non-reducible kind, *consciously judging that* p.

[25] Discussions of the relation between thinking and inner speech go back at least as far as Plato. In the *Theaetetus* he has Socrates saying, 'I have the notion that, when the mind is thinking, it is simply talking to itself, asking questions and answering them . . . So I should describe thinking as discourse and a judgement as a statement pronounced, not aloud to someone else, but silently to oneself.' Note the argument of this chapter does not purport to show that the occurrence of inner speech, as opposed to other forms of phenomenally conscious mental event, is necessary for conscious judging. For the idea of understanding judging as the interiorization of an act of asserting, rather than understanding asserting as the exteriorization of an act of judging, see Dummett 1973: 362. See also Geach 1957; Kenny 1963: ch. 10; Sellars 1963: s. 50.

IV

To help clarify the account I'm suggesting I'll outline and discuss a potential objection to it that is based on a development of Budd's reading of some remarks of Wittgenstein on inner speech in the *Philosophical Investigations*.[26] The tenor of the objection is that, although it may be tempting to model calculation in one's head on calculation out loud, the account of the 'inner' process one thereby ends up with is ultimately untenable.

Calculation out loud or on paper requires a vehicle: the concept demands that the calculation is embodied in a process that begins when the subject begins to calculate and terminates when he reaches the conclusion. The process has an independently specifiable nature as well as satisfying a description in terms of what is calculated. There are similarities here with calculation in the head: the concept is of something that occurs within a stretch of time and that can be said to consist of stages or steps. This encourages the following picture. In the case of calculation in the head the role of the vehicle of the calculation is played by an inner process, inner speech (which has an independently specifiable nature), rather than an outward process.[27] In the case of the outward process, there is room for mistake concerning what is happening, but there is no room for error or ignorance about my inner speech. So in modelling calculation in the head on calculation out loud, we are tempted into accepting that when I calculate in my head:

(*a*) something happens in me
(*b*) which I bring about,
(*c*) which is the internal analogue of what happens and is brought about by me when I speak or calculate overtly,
(*d*) which I have direct and infallible concurrent awareness of.

But there are reasons why we should not be seduced by this picture, which looks very much as though it commits to the Cartesian Theatre model of self-knowledge. First, in order to judge that *p* it is unnecessary that I should produce imagined sounds in my head that only I hear. Although the act of assertion involves a vehicle of representation *and communication*, the mental act of judging need not involve a vehicle of communication. The act of assertion seems to require the *production* of something that acts as the vehicle of communication. In contrast, the mental act of judging does not seem to require the production of anything.[28] This is why although it makes sense to ascribe to the agent an

[26] See Budd 1989. [27] See Wittgenstein 1953: 366.

[28] Compare O'Shaughnessy 2000: 'There exist active mental phenomena which at first glance may look to be the immediate producing of a desired internal event, but which on inspection prove to be the active producing of *nothing*: sheer doings as one might say. The best example of this is

intention to assert that *p*, it doesn't make sense to attribute to the agent an intention to judge that *p*. One way of putting this is to say that the mental act of judging is not to be understood as inner assertion, where this is a matter of communicating something to oneself. I do not need to produce inner sounds that I hear with my inner ear in order to understand and know what I am judging.

Since the mental act of judging, in contrast to the act of assertion, does not require the production of a vehicle of communication (that one needs to hear with one's inner ear in order to know what one judges), then it seems odd to think of the mental act of judging as requiring the production of anything. In which case, what use have we for the notion that the mental act of judging involves a phenomenally conscious mental act that serves as a *vehicle* of representation? For the notion of vehicle in play here seems to be the notion of an act or action one performs (or something produced by an act or action one performs) in order to represent or signify something—the notion that there is some vehicle one makes use of in order to signify/represent something. Which in turn suggests that there is someone (onself?) to whom one represents things by use of that vehicle.

Similar concerns about making too much of the analogy between speech and thought are expressed in the following remarks by Vendler:

> Whereas speech is the expression of thought in a code—that is by means of a language—thought is not an expression of anything and is not conceived in or via a code. It is inconceivable that I might fail to understand what I think. Hearing the speaker's voice, or seeing his writing, is indispensable for getting at what he said, but what do I have to see or hear, externally or in my mind, to get at my own thoughts? [In the case of thinking] there is no message to encode and no private language to use for the encoding. (1972: 42)

> Encoding and decoding can be correct or incorrect—misunderstanding, and slips of the tongue etc. are possible. If thinking needed a code, consisting of words or other symbols, then on the one hand the thinker might know what he wanted to 'say' to himself, and on the other he could be mistaken about what he did 'say'. In other words it would be possible for him to know and not know what he thinks at the same time. This is absurd . . . (1972: 44)

Considerations of this kind lead Vendler to hold that mental acts of thinking thoughts are not to be identified with 'the flux of words and images we perceive with the imagination'. Geach, in a similar vein, rejects the idea that 'to think certain thoughts is to have certain mental images, feelings, unspoken words etc., passing through one's mind' (1969: 34).[29]

silently talking to oneself. We do not need to internally hear in order to internally speak. Is silent speech the production of imagined sounds? Why should success be of any acount to the subject? Why should it matter to the silent speaker that seeming sounds *actually* issue forth in his own mind?' (p. 380)

[29] For an argument against the claim that we engage in purely propositional thinking, see Carruthers 1998: 118: 'I propose . . . that what we often describe as purely propositional (non-verbal) thoughts, available to introspection (and hence conscious), are really the results of active

The correct response to this line of objection is to concede that as the agential mental activity of thinking (e.g. calculating whether *p*) differs in a number of ways from the agential activity of communicating, different constraints operate on what counts as performing these actions. The activity of asserting is constitutively linked with the notion of communication in a way that the mental act of judging is not. What we need to do is make sense of how it is that we can depart from the implicit assumption that the only sense that can be made of the claim that the mental act of judging requires a phenomenally conscious mental act to serve as a vehicle of representation, is to think of that phenomenally conscious mental act as serving as a vehicle of communication—the notion that one makes use of the vehicle in order to represent something to someone. For it is this assumption that makes trouble for the idea that the conscious act of judging requires a phenomenally conscious mental event to serve as vehicle of representation.

In the case of the act of judging, the phenomenally conscious vehicle of representation is not a vehicle of communication. One does not make use of a vehicle in order to represent something to someone. In contrast to the act of assertion, the mental act of judging does not involve any such aim.[30] So it is a mistake to think that when one consciously judges that *p* one brings about/produces a mental event with the aim of *thereby* representing that *p* to someone. This is what is correct in the claim that the act of judging does not involve the production of anything. It is a sheer doing.

So what then is the role of a vehicle of representation in the case of conscious judging, if it is not that of serving as a vehicle of communication? Let us return to the question of what initially led to the thought that the mental acts involved in conscious thinking are mental events with duration. It was the suggestion that the mental acts involved in conscious thinking must have duration if conscious

self-interpretation. So even when the interpretation in question happens to be correct, and the thoughts are self-ascribed veridically, these thoughts are *not* conscious ones. Compare also Burge, who writes, 'A being that lacks phenomenal consciousness could not be conscious in any way. It would not, for example, have imageless conscious thought' (1997: 429; see also O'Shaughnessy 2003: 355).

[30] Dummett, who suggests that we should understand judging as the interiorization of an act of asserting rather than understanding asserting as the exteriorization of an act of judging, writes, 'The reason why there can be an interior analogue of the conventional act of assertion is that the use of language is not of purely social significance . . . The possession of language alters the behaviour of an individual quite apart from his immediate dealings with others . . . The use of language for private, rather than social, purposes can be interiorised: the result is the activity of judgement' (1973: 363). Compare also Sellars's (1963) discussion of Jones, our 'Rylean ancestor', who introduced the concept of thought that applies to inner episodes of thinking. According to this myth, in the attempt to account for the fact that his fellow men behave intelligently, not just when they 'think out loud', but also when no detectable verbal behaviour is present, Jones develops a theory according to which overt utterances are the culmination of a process which begins with certain inner episodes. Significant here is Sellars's suggestion that the model for these inner episodes is overt verbal behaviour, together with his claim that, although the theory involves a model, it is not identical to it.

thinking is to satisfy the kind of accessibility requirement necessary for mental action. So implicit here has been the suggestion that we need to think of the kind of mental events involved in conscious thinking as such that a sequence of these events can amount to a mental activity that has an *agential* explanation. These events must be such that appropriate sequences of them can instantiate mental actions.

The kind of events involved in conscious thinking must be such that an appropriate sequence of them can constitute activity that is attributable to the agent, as opposed to some part of him. The activity of conscious calculation constitutively depends on the obtaining of certain relations between the mental acts that make up that activity, and the obtaining of those relations, in turn, depends upon the way in which the contents of the mental acts are structured. In order for the activity of conscious calculation to be attributable to the agent, as opposed to some part of him, the agent must himself have some grasp of the way in which the contents of the mental acts that make up that activity are structured. The agent's use of vehicles of representation is what manifests his knowledge of the way in which the representations of which they are vehicles are structured. So the notion of a vehicle of representation that the agent makes use of is something we need in play if we are to make sense of the occurrence of conscious calculation as an activity that is attributable to the agent—an activity that the agent can be intentionally engaged in.

The proposal here then is that we do not get at the notion of mental action through adding to the notion of conscious mental act, rather we get at the notion of conscious mental act by subtracting from the notion of mental action. Once we acknowledge that we get at the notion of a conscious mental act by subtracting from the notion of mental action, and that such conscious mental acts must involve vehicles of representation if appropriate sequences of them are to constitute mental activity that has an agential explanation, then we can make sense of the idea that there can be events with duration, and so which are not achievements, which can be correctly described as vehicles of representation of acts that are achievements.

A further objection to the view being outlined here may target the kind of account of self-knowledge that it assumes. When one is, say, consciously calculating whether *p*, how does one know what one is doing when one consciously judges that *p*? Aren't proponents of this model committed to a version of the much-criticized Cartesian Theatre model of self-knowledge? On the present picture, doesn't one have to observe and interpret a phenomenally conscious mental event—e.g. one's saying something in inner speech—in order to know what thinks? For otherwise, can't one know what one is doing in consciously judging that *p* without having to actually say anything in inner speech? Consider again the case of judging that *p* out loud. Here it seems that the following counterfactual is true: I would have known that I was judging that *p* even if I failed to produce any sounds at all. So shouldn't we say that in the

case of consciously judging that *p*, I would have known that I was judging that *p*, even if I hadn't actually said anything in inner speech?

One might argue that the event of acquisition of one's knowledge of the content of one's mental act of judging must be instantaneous. And if the event of acquisition of one's knowledge is instantaneous then the unfolding of the phenomenally conscious mental event that is supposed to be the vehicle of one's conscious mental act of judging is either redundant, or what one 'observes' and interprets in order to acquire the knowledge. If one's silently saying that *p* is supposed to be the conscious vehicle of one's mental act of judging that *p*, then either one already knows what one is doing as one starts silently saying that *p*, or one must silently say that *p* in order to know what one is doing. The former option suggests that the unfolding of the phenomenally conscious event of one's silently saying that *p* is inessential to one's consciously judging that *p*, since it suggests that one can know what one is doing in judging that *p* without having to silently say that *p*. And the latter option suggests that one must wait for the completion of this phenomenally conscious event, and interpret it, in order to know what one thinks.

The solution, I suggest, is to regard the phenomenally conscious mental event that unfolds over time, and which is the vehicle of one's consciously judging that *p*, as an event that *manifests* one's knowledge of the content of one's mental act—a mental act that is an achievement. Without the occurrence of a phenomenally conscious mental event that unfolds over time and that manifests one's knowledge of the content of one's mental act, there is no belief that one is performing such a mental act. At most, there is simply a second-order mental state with another mental state (or process) as its object, and not a mental act; or there may be a belief that one has just performed a mental act—a belief that might be manifested in an out-loud verbal utterance—a report about what one thinks one has just done.

The occurrence of a phenomenally conscious mental event that is the vehicle of an act of conscious judging only manifests one's knowledge of what one is doing if it can be thought of as manifesting a mental state that plays a particular kind of role in the mental life of the subject.[31] And on this view, one does not, as the so-called Cartesian model suggests, acquire knowledge of what one is doing by observing and interpreting the occurrence of these phenomenally conscious mental events. Self-awareness is not here being modelled on other-awareness. Rather, the phenomenally conscious mental events involved in cases of conscious judging should be thought of as events that *manifest* self-knowledge—one's knowledge of what one is thus doing.

[31] Compare here Ewing, who writes, 'With cognition . . . we should distinguish a continuous process of thinking from particular "cognitive acts". The former should, I think, be regarded as basic rather than the latter, and the process of thinking out a problem should not be reduced to a mere series of such acts' (1948: 217).

I have been exploring the proposal that the difference between our third-person access to, on the one hand, overt bodily actions and, on the other, the kind of covert mental action I have been focusing on, can affect how we conceive of that variety of mental activity. The idea that the mental events that constitute such mental actions are not third-personally observable may suggest that we are forced to choose between, on the one hand, a Cartesian model of our own private and infallible access to such entities, or, on the other hand, the conception of such actions as simply being a variety of mental process that involves transitions between mental states and that enables and explains certain overt bodily actions. The former option seems untenable, and the latter option suggests that mental actions are simply to be thought of as mental processes that play a certain kind of functional role, which in turn suggests that they need not involve the occurrence of any phenomenally conscious mental events.

Although their phenomenal characters do not individuate many of the mental actions I have been considering, it is a mistake to conclude from this that such mental actions can be carried out without the occurrence of any phenomenally conscious mental events.[32] The claim that the mental actions I have been focusing on must involve the occurrence of phenomenally conscious mental events, was reached *via* a consideration of the kind of access an agent must have to such mental activities if they are to be regarded as properly agential. But this need not involve a commitment to an objectionable 'Cartesian Theatre' model of the mind. I do want to suggest, however, that there is a Cartesian insight that should be preserved, and I want to close by contrasting certain features of Descartes's conception of mind with more recent approaches.

V

As is well known, in current philosophical debates it is suggested that certain conscious aspects of mind present the 'hard' problem for physicalism.[33] Descartes presumably thought that there is something about certain *conscious* aspects of

[32] The propositional content of the mental act of judging is not determined by the phenomenal character of the mental act that is its vehicle. That is, mental acts with different phenomenal characters can act as vehicles of acts of judging that have the same contents and mental acts with the same phenomenal character can act as vehicles for acts of judging that have different contents. However, note that this is not to say that the phenomenal character of the mental act that is its vehicle exhausts the phenomenology of an act of judging. For knowledge of what one is doing (e.g. knowledge that one is judging that *p*) may affect the phenomenology of so acting. A further point is that if the occurrence of phenomenally conscious mental events is crucial to the performance of certain mental actions that are not individuated by their phenomenal characters (e.g. calculating and deliberating), it seems to reasonable to expect that a creature capable of engaging in such actions will also be capable of engaging in mental actions that *are* constrained, and so individuated, by their phenomenal characters (e.g. varieties of sensory imagining and perceptual memory).

[33] See Chalmers 1996.

mind that present the 'hard' problem for materialism. Indeed, Descartes is often criticized for focusing exclusively on conscious aspects of mind. But it is noteworthy that the conscious aspects of mind that philosophers now tend to suggest create the hard problem for physicalism are not the aspects of mind Descartes focused on, and the problems that they purport to raise are not the same as the ones he raised.

A major and obvious difference between the Cartesian approach and the contemporary approach is that Descartes was interested in the possibility of disembodied existence in a way that much of the current debate is not. The very idea of disembodied existence seems to unravel if we conceive of mental actions as simply processes that enable and explain overt bodily actions.[34] So Descartes obviously had a motivation to reject such a conception of mental action. It seems that we cannot make sense of the notion of disembodied existence if we cannot makes sense of the notion of disembodied agency, and furthermore, it seems that we cannot make sense of disembodied agency without thinking of the disembodied agent as capable of engaging in *conscious mental action*. For regarding mental activity as agential simply in virtue of its association with agential, overt bodily action is obviously not an option if one is considering the mental life of a disembodied agent.[35]

Since much contemporary philosophy of mind does not take seriously the possibility of disembodied existence, there is not the same motivation for denying that mental action is to be conceived of as simply a variety of mental process that involves transitions between mental states that enable and explain overt bodily action. And for the reasons that I've outlined, it can be tempting to think that there can be conscious mental action without the occurrence of any phenomenally conscious mental events. The upshot is that many recent theories do not think of mental events with phenomenal properties as things that must be rationalized by mental state attributions, or as *mental behaviour* that is the *output* of the functional roles that mental states characteristically play. They are either inputs (e.g. sensations and experiences) or intermediaries (i.e. mental events constituting processes that are themselves individuated by their functional roles).

So whereas for Descartes it is the fact that conscious thinking is not simply to be thought of as a process that enables and explains bodily action that purports to present the hard problem for materialism, for the contemporary philosophers it is the fact that phenomenally conscious mental events are to be thought of as inputs and intermediaries that presents the hard problem for physicalism. For then the functional role of the phenomenal properties of these mental events becomes problematic.

[34] See Shoemaker 1976.

[35] It is, I think, significant that the scepticism of Descartes's *First Meditation* targets one's knowledge of one's own *bodily actions*—e.g. shaking one's head, stretching one's hand out deliberately.

I suggest that the Cartesian insight we need to reinstate in order to approach these issues with the appropriate conceptual framework involves taking seriously the notion of *agential* mental activity that is not simply to be thought of as a variety of mental process that involves transitions between mental states and that enables and explains overt bodily action. A consequence of reinstating what I am calling the Cartesian insight is that various assumptions that usually frame debates about phenomenal consciousness may be undermined.[36]

Since we *are* embodied, we *can* think of our mental actions simply as mental processes involving transitions between mental states that enable and explain our overt bodily actions. But we overlook the Cartesian insight if we conceptualize such actions in this way, and we may thereby end up creating conceptual obstacles to a proper understanding of phenomenal consciousness.[37]

REFERENCES

Audi, R. (1999) 'Doxastic Voluntarism and the Ethics of Belief', *Facta Philosophica,* 1: 87–109.

Budd, M. (1989) *Wittgenstein's Philosophy of Psychology* (London: Routledge).

Burge, T. (1997) 'Two Kinds of Consciousness', in Ned Block, Owen Flanagan, and Güven Guzeldere (eds.), *The Nature of Consciousness* (Cambridge, MA: MIT Press), 427–33.

Byrne, A. (2004) 'What Phenomenal Consciousness is Like', in R. Gennaro (ed.), *Higher-Order Theories of Consciousness: An Anthology* (Amsterdam: John Benjamins), 203–25.

Carruthers, P. (1998) 'Conscious Thinking: Language or Elimination', *Mind and Language,* 13: 457–76.

——(2000) *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press).

——(2006) 'Conscious Experience versus Conscious Thought', in U. Kriegel and K. Williford (eds.), *Consciousness and Self-Reference* (Cambridge, MA: MIT Press), 299–320.

Chalmers, D. (1996) *The Conscious Mind* (Oxford: Oxford University Press).

Davidson, D. (1971) 'Agency', in his *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

Dennett, D. (1969) *Content and Consciouness* (London: Routledge & Kegan Paul).

[36] Compare here Burge, who writes, 'lacking some . . . frame of phenomenal consciousness one's thinking could not be conscious in any way. A phenomenal zombie has no consciousness—no matter how efficiently rational its behaviour, verbalizations and reasoning. I do not know how to defend this view. I do not know why it is true. But despite a literature replete with assumptions to the contrary, I find it compelling' (1997: 429).

[37] I am grateful to Lucy O'Brien, Guy Longworth, and two anonymous referees for very helpful comments.

——(1991) *Consciousness Explained* (Boston, MA: Little, Brown).
Descartes, R. (1996) *Meditations on First Philosophy*, tr. J. Cottingham (Cambridge: Cambridge University Press).
Dummett, M. (1973) *Frege: The Philosophy of Language* (London: Duckworth).
Ewing, A. C. (1948) 'Mental Acts', *Mind,* 57: 201–20.
Flanagan, O. (2000) *Dreaming Souls: Sleep, Dreams and the Evolution of the Conscious Mind* (Oxford: Oxford University Press).
Fodor, J. (2001) *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* (Bradford, PA: MIT Press).
Geach, P. (1957) *Mental Acts* (London: Routledge & Kegan Paul).
Ginnane, W. G. (1960) 'Thoughts', *Mind*, 69: 372–90.
Hornsby, J. (1997) *Simple Mindedness: In Defence of Naïve Naturalism in the Philosophy of Mind* (Cambridge, MA: Harvard University Press).
Kenny, A. (1963) *Action, Emotion and Will* (London: Routledge & Kegan Paul).
Lear, J. (1990) *Love and its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis* (New Haven, CT: Yale University Press).
Marcel, A. (2003) 'The Sense of Agency: Awareness and Ownership of Action', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press).
Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press).
Mourelatos, A. P. D. (1978) 'Events, Processes and States', *Linguistics and Philosophy*, 2: 415–34.
Mouton, D. M. (1969) 'Thinking and Time', *Mind,* 78: 60–76.
O'Brien, L. (2003) 'On Knowing One's Own Actions', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press).
O'Shaughnessy, B. (1974) 'Trying as the Mental Pineal Gland', *Journal of Philosophy,* 70: 365–86.
——(2000) *Consciousness and the World* (Oxford: Oxford University Press).
——(2003) 'The Epistemology of Physical Action', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press), 345–57.
Peacocke, C. (2003) 'Action: Awareness, Ownership and Knowledge', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness* (Oxford: Oxford University Press).
Plato (1961) *Theaetetus*, in *The Collected Dialogues of Plato,* ed. E. Hamilton and H. Cairns (London: Pantheon).
Roessler, J., and N. Eilan (eds.) (2003) *Agency and Self-Awareness* (Oxford: Oxford University Press).
Ruben, D. H. (2003) *Action and its Explanation* (Oxford: Oxford University Press).
Sellars, W. (1963) 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality* (London: Routledge & Kegan Paul), 127–96.
Shoemaker, S. (1976) 'Embodiment and Behavior', repr. in his *Identity Cause and Mind* (Oxford: Oxford University Press).
Steward, H. (1997) *The Ontology of Mind: Events, Processes and States* (Oxford: Oxford University Press).
Vendler, Z. (1957) 'Verbs and Times', *Philosophical Review*, 66: 143–60.

Vendler, Z. (1972) *Res Cogitans: An Essay in Rational Psychology* (London: Cornell University Press).

Walton, K. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Cambridge, MA: Harvard University Press).

Wittgenstein, L. (1953) *Philosophical Investigations,* tr. G. E. M. Anscombe (Oxford: Basil Blackwell).

13

Is there a Sense of Agency for Thought?

Joëlle Proust

INTRODUCTION

Are we acting when we think? When your body moves, there is a sense in which one may ask whether you moved intentionally, or whether someone pushed you, for instance. However, there is no consensus about there being any equivalent possibility with thoughts. Thinking encompasses all sorts of different attitudes, from considering, judging, comparing, evaluating, and reasoning to imagining, visualizing, desiring, intending, planning, and deciding. Although it is uncontroversial that each thinker has a specific, privileged connection to her own thoughts and thought processes, many philosophers agree that thought contents are determined by the environment rather than by the thinker.[1] One often has, however, a distinctive impression when one is thinking a thought, whatever its content: there is a sense of being the thinker of that thought, and a sense of this thought as being the one that is presently occupying one's attention. In the recent literature this sense is called the sense of 'owning' a thought—of having first-person knowledge of one's having this thought.[2] One speaks of the sense of 'ownership' or of 'subjectivity' in thought by analogy with the experience of acting, where an agent can feel her body involved in a willful movement. The second—more contentious—type of experience associated with thinking is that of intending to think this particular thought: it is the sense of feeling agentive while thinking. It is called 'the sense of agency', by analogy again with the awareness of action; to feel active in an action is an experience that differs from the sense of having a characteristic bodily experience while acting.

Although thinking is a mental activity that has a purpose, uses resources, may need time to be completed, etc., is it a willful activity? Here intuitions seem to diverge considerably. One common way of addressing the question consists in asking: does it *'feel'* willful? Answers are surprisingly varied.

[1] See e.g. Burge (1998*b*).

[2] See Stephens and Graham (2000); Campbell (2002); Proust (2006*a*).

Some people see their successive thoughts as something they are acting upon in their contents and even in their formal relations. They see their own thoughts as the expression of their rationality, and of their own self; they insist that thinking involves commitments to epistemic and moral values such as truth, responsibility, and dependability. Others, however, take thinking to occur mostly outside of awareness. Beliefs and desires occur to us; reasoning does not seem to leave room for choices or stylistic variations. Thoughts seem sometimes to be entertained and to determine our behaviors with no associated subjective awareness, let alone any sense of agency.

The challenge is made still more pressing by the fact that psychopathology offers additional puzzles in this area. The very distinction between a sense of agency and a sense of ownership was introduced in the philosophical literature to account for the subjective experience of many deluded patients with schizophrenia.[3] Although they have normal proprioceptive and visual experience while acting (and therefore, a preserved sense of ownership), they often feel that someone else is acting through them[4] (they present a disturbed sense of agency). Another frequent delusion, however, is still more intimately associated with self-knowledge: patients experience 'thought insertion'; they complain that some of their thoughts are in their minds (and, to this extent, are experienced subjectively), but at the same time are not theirs in the agentive sense; they speculate retrospectively that someone else has inserted them 'into their heads', making them think these ideas (using futurist brain technology, or otherwise).[5] Interestingly, these patients also have the impression that some or most of their intentions to act are not theirs. They feel that someone else is willing them to act the way they do.

One simple way to interpret this symptom is the following. Patients affected with thought insertion teach us that one can lose one's sense of agency for thoughts as one can for willful movements. If one can lose it, suddenly realizing that one is having only, or mainly, passive thoughts, then it should be recognized that this is a conscious feature that one has had all along. This interpretation is contentious, however, for a sceptic might argue that two other possibilities are still open. The first is that the patient may be correct when vividly sensing what one normally does not sense but may only infer: that the thoughts he is having or has just had are mostly *not* under his own control, that they really are/were 'inserted' into his mind.[6] After all this might be how our beliefs are

[3] See Daprati *et al.* (1997); Frith *et al.* (2000); Farrer and Frith (2002); Farrer *et al.* (2003).

[4] In the sense that foreign intentions, rather than their own, appear to them to be causing their behavior.

[5] This parallel between perturbations in thought and action suggests a common explanation in terms of a specific control deficit. See Proust (2006*a*).

[6] We must ignore in the present discussion an additional problem raised by the schizophrenic delusion of control, namely the feeling of 'insertion' or 'external control', i.e. the attribution of thought agency to another agent by the deluded thinker. Failing to experience agency for one's

formed: automatically, inevitably, and mostly or even exclusively under external influence.[7] If the sceptic[8] is right, *normal subjects* would generally be *wrong* in attributing to themselves agency in thought. Thoughts are the moment-by-moment expression of an unending process of combination and retrieval; they exploit brain structures, inferential principles, and motivations of the system in much the same way as viruses do;[9] they don't engage any 'authorship' of a thinker.

A second, stronger claim would be that the so-called 'senses' of agency as well as of passivity in thought might actually *both* be lacking in normal thinkers; when a sense of passivity in thought is felt, it would then be an experience of a hallucinatory kind, for there would actually be nothing to be sensed at all, no information channel allowing one to detect the purportedly active or passive ideas. If one is hallucinating, one may wrongly believe that one is active, or that one is passive, but neither belief would be true.[10]

As a consequence of this strong disagreement with respect to the phenomenology of thought, we cannot take subjects' reports at face value to claim that there is a sound analogy between thinking and acting. One should insist that a sense of agency is veridical only for those occurrent thoughts, if any, which are under our will, namely those that independently qualify as mental actions. If no thought can be willed or tried, no sense of agency in thought should surface. But merely invoking a feeling of agency does not seem to be a promising route for rejecting the sceptical considerations.

This disagreement has significant epistemological consequences. There are two ways of characterizing the epistemic rights of a believer: explicit justification or implicit entitlement.[11] One's belief is justified if one knows what reasons one has

thought does not automatically generate a sense of being acted through; an additional, projective, component is present in the control delusion, and absent from the common phenomenology of thinking thoughts. What explains this difference? Possible solutions may point to content properties (attributions are driven by those contents that contradict the subjects' beliefs and motivations), to functional properties (by those attitudes that are intrinsically agentive, such as intentions), or to structural properties (by the neural vehicles, such as the inferior parietal lobule, forcing an extraneous attribution in a non-systematic, contingent way).

[7] See Williams (1971) and Strawson (2003) for detailed expositions and discussions of this claim.

[8] Galen Strawson reflects the skeptic position sketched above, when he writes: 'Those who take it, perhaps very unreflectively, that much or most of their thinking is a matter of action are I believe entirely deluded' (2003: III).

[9] On the memetic view of representations: see Dawkins (1976); on an alternative 'epidemiological' view of representations and belief fixation, see Sperber (1996) and Boyer (2002).

[10] This line of argument raises the problem of the conceptual and empirical soundness of the view that a hallucinatory form of experience supervenes on no correct variety: visual hallucinations do depend on visual perception. Similarly agency-in-thought-hallucinations should presuppose that one can normally perceive agency in thought. But one might argue that the patient hallucinates a sense of agency in thought on the basis of his/her normal sense of agency in action, in particular in speech.

[11] See Sosa (1991). For a full defense of a moderately externalist view, see Dretske (2000).

for believing that *P*. One is entitled to believe that *P* when one's experience is so compelling that one can only form the belief that *P*, and one has no reasons to mistrust the way this belief was formed. If the disagreement about active thinking prevailed, we would have to say that a subject may never be justified, or even entitled, to know when, and even whether, she acts mentally.

We will examine below two types of theories about thinking as a matter of agency that do try to provide a theory of mental action that responds to the sceptic's worries. The first is a theory that originates in a strict analogy between thinking and acting: thinking is a type of bodily action, in that it is a covert type of motor activity, and engages substantially the same kind of mechanisms. We will see that this theory presents insuperable problems. However, it contains interesting ideas about action as a control structure, which we will retain in our own proposal.

A second type of theory aims to identify mental actions based on their 'trying' structure, a structure that is supposed to apply across the bodily and mental realms alike (sections 2 and 3). It avoids the sceptic's claim by considering that the form of action-awareness involved in thinking does not automatically produce or justify a corresponding belief to the effect that such and such a mental action was performed, or was performed successfully. Although this theory brings into focus important distinctions, it is still incomplete in significant respects. A fourth section of this chapter will attempt to provide needed complements through a control view of mental action. This definition will help us determine a functional connection between agency in thought and metacognitive thinking. Section 5 will finally defend the view that metacognitive feelings are the ground which entitles a thinker to form predictions and evaluations based on her mental actions. We will apply this view to the feeling of having inserted thoughts in schizophrenic delusions.

1. THE MOTOR THEORY OF THINKING

The neurosurgeon Wilder Penfield used to conduct experiments on the open cortex of his patients while they were awake to test their subjective experience. He found that subjects caused to move by stimulation of their motor cortex would deny agency for that movement. Such a response is a good illustration of what a control view of the motor system would predict. According to this view, a subject has a feeling of agency for her movements when she is in a position to anticipate and evaluate the consequences, both internal and external, that are associated with them.[12] Another experiment by Penfield suggested a generalization of

[12] At first, it was hypothesized that the cancellation of a motor command thanks to a corollary discharge was both the source of the awareness of agency and instrumental for identifying the relevant feedback (Sperry 1950; von Holst and Mittelstaedt 1950). MacKay (1966), however,

this account to thinking: subjects 'made to remember' by stimulation of their temporal lobes would also report a sense of externality: 'you caused me to think that.'[13] A natural supposition is that, in both cases, a motor process is involved (with its associated command and anticipated feedback). Accurately predicting the total feedback for a given sequence labels the upcoming thought or movement as being internally generated. This generalization was prepared for by Hughlings Jackson's popular view[14] that mental operations exploit the organization of the sensorimotor system. Under Jackson's recognized authority, the psychiatrist Irwin Feinberg posited the following conditional statement: 'If thought is a motor process heavily dependent upon internal feedback, derangement of such feedback might account for many of the puzzling psychopathological features of the "psychosis of thinking" [thought insertion]'.[15] Penfield's observation of the patient 'made to remember' is phenomenologically important. It seems to add credit to the view that patients with schizophrenia have a disturbed sense of agency for thought. Let us summarize, however, the difficulties with Feinberg's speculation.

It is far from clear that a predictive ability makes any sense in the case of thinking. It is doubtful that a central motor command is used by the brain to keep track of its remembering activity. The feeling of willful activity that is felt by a normal subject, and is missing in Penfield's patient, might be inferred rather than directly perceived. The patient who abruptly remembers, out of context, a specific memory, might reject agency on the basis of a lack of continuity with his stream of thought; while a thinker engaged in natural remembering might accept agency because of the redundant properties of his stream of thought. This does not show that motor activity took place, nor even that there was a mental act. The 'natural' subject may simply have inferred on the basis of the occurrent context of his thinking that his thought content was of an expected type: having a memory, say, rather than a sudden burst of planning or a desire for a mountain hike. If these objections to the motor account are correct, however, thought agency might dissolve into thought ownership: introspective aspects of thought content, such as ease of access, familiarity, redundancy, might have more to do with having a subjective feeling that one has a thought (ownership)

showed that the concept of feedforward control provides a better account for the brain's ability to monitor reafferences through feedback prediction. This kind of model is now widely used in understanding both action and action awareness in the normal and in the deluded subject. See Proust (2006*a*, 2007). On this view, a 'dynamic model' of the type of action has first to be used to select the specific command leading to the required result; secondly, an 'efferent' copy of the command must be available throughout the action to allow the subject to identify the relevant feedback for *this* action.

[13] Penfield (1974). [14] Jackson (1958).

[15] Feinberg (1978: 638). Notice however that Feinberg recognized that he had no independent evidence in favor of his premise. What he was offering was an evolutionary speculation rather than an empirically established claim that thought insertion is caused by the absence of a 'corollary discharge associated with the motor act of thoughts'.

rather than with agency (the feeling that one is 'deliberately' producing that thought).

Because he had similar problems with Feinberg's motor theory of thinking (1998, 1999, 2000), John Campbell attempted to revise it in order to get a clearer distinction between ownership and agency in thought. Campbell retains the gist of Feinberg's proposal: in schizophrenia, the preserved sense of ownership in thought is dependent on 'introspective knowledge', whereas the disturbed sense of agency is modulated by a mechanism allowing self-prediction (similar to the efferent copy of action commands). Campbell's proposal uses a belief–desire causation of action framework to rescue the motor view of thought. A motor command is needed to activate each token of thought: 'the background beliefs and desires cause the motor instruction to be issued', which 'causes the occurrent thought' (Campbell 1999: 617). This explains 'how the ongoing stream of occurrent thoughts can be monitored and kept on track' (ibid.). Campbell's 'motor' view of thinking thus relies on the plausible intuitions that thinking consists of inner speech, and that inner speech is closely related to outer speech: they are equally serial, they have roughly the same duration, and they share resources (it is difficult to simultaneously say one thing out loud and another silently to oneself).[16] Given that speech engages motor activity, it might well be that speaking to oneself also does.

This proposal is not fully convincing, however. First, it is not clear that background mental states must cause a motor instruction in order to cause a thought. There is evidence that trying to imagine oneself walking—or performing any other bodily action—activates a (pre) motor instruction as well as the corresponding thought of what it is like to walk.[17] But most cases of thinking do not include any reference to an action, and thus cannot automatically activate *motor* representations. It seems implausible, prima facie, that symbol activation and sentence generation 'in the head' actually involve 'manipulating' items, which would in turn require motor command and efference copy. Nor does silent speech seem necessarily involved in all kinds of thinking: spatial thinking and visualizing, for example, do not seem to necessarily or always rely on words. So even though the motor hypothesis is able to account for a category of thinking episodes—which might turn out to be causally relevant for thought insertion—it cannot provide a general explanation of how thinking develops, and of how a thinker gets the sense of acting-in-thinking.

A second problem is that many thoughts come to mind without a prior intention (or even without any 'intention in action') that would put the current ideation under immediate intentional control. Indeed if every thought presupposed a former intention, and if such an intention is a form

[16] Lormand (1996: 246). [17] See e.g. Blakemore and Decety (2001).

of thinking, we would seem to have an infinite regress.[18] It does not seem, however, that we normally *intend* to move from one thought to the next. The process of thinking does not seem to be constrained, in general, by former intentions.

Finally, the motor theory is too strong, in that it should lead a normal subject to acknowledge thought agency independently of her thought contents. Many of our thoughts, however, are *not* experienced as fully ours; for example, in a conversation, we process thoughts that are conveyed to us, and that we may imperfectly grasp; we have no trouble both having the sense that we entertain a thought, understand it in part, process its consequences etc., and attributing its source to another thinker. This mechanism of deferential thinking is fundamental as an early step in belief fixation.[19] Any theory of agency in thought should be able to account for the *degree* to which a thinking episode is perceived as agentive by the thinker.

To summarize, Campbell's 'motor' hypothesis is valuable in delineating the various dimensions involved in the problem of thought insertion as an exception to immunity to error through misidentification.[20] The motor view, however, seems to have a clearer meaning in the case of bodily action than in the case of thinking. Even if one concedes that the motor view can indeed account for a subcategory of thinking, requiring intentional inner speech, a category that would be causally involved in thought insertion phenomena, it does not offer a general explanation of agency in thought. Firmer ground is needed for Jackson's claim of a functional similarity between acting and thinking.[21]

2. ACTIVE THINKING AS MADE FOR REASONS AND ANSWERABLE TO REASON

Most philosophers who have recently explored the domain of mental action have done so to underscore the difference between automatic belief fixation and belief acquired through critical reasoning. Their view on mental action is relevant in the present perspective, for it is an easy step to generalize from the epistemic

[18] This objection was articulated in Gallagher (2000). See Ryle (1949) for a general presentation of this argument. Even if one accepts the view that intentions are acts of thinking, as we do here, the objection can be disposed of. See Proust (2001).

[19] On deference, see Recanati (2000).

[20] Campbell's view suggests that (but without explaining how and why) a loss of agency in thought—with preserved introspection—should lead to a disintegration of the concept of self: 'the very idea of a unitary person would begin to disintegrate if we supposed that thoughts were generated by people other than those who had introspective knowledge of them' (Campbell, 2002: 36). What exactly does 'disintegration' mean? Is the concept of self lost, or occasionally misapplied? And how can introspection be preserved in case one's thoughts are not sensed as one's own?

[21] For a similar diagnosis, see Gallagher (2000); Gerrans (2001); Spence (2001).

to the general case. To introduce this view, it will be easier to start with L. J. Cohen's distinction between disposition and mental 'act'.[22]

In *Belief and Acceptance*, L. Jonathan Cohen distinguishes the disposition to believe (as the disposition to 'credally feel' that *P*) from the 'mental act' or 'policy' of accepting *P*, which involves the commitment to use *P* as a premise in reasoning and decision-making.[23] The importance of this opposition is that you can willfully and deliberately accept as a premise a proposition that you don't believe to be true, i.e. on a prudential rather than on an evidential basis. There is a pragmatic dimension involved in premising that can overrule belief, a dimension that seems to require an active and explicit decision—a 'policy'—from the thinker.

In some cases, judging may result from a shallow form of believing: registering facts delivered by perception, inference, or testimony. But in other cases, in its critical usage, 'judging' expresses the *decision* to use a certain representation as a premise, more or less independently of the evidence that supports it (evidence can be overruled by prudential considerations). A speaker may also choose which premises to select as a function of her audience. These observations suggest that forming (or expressing) a judgment is sensitive to context, can be a topic for deliberation, and can be governed by prior intentions. Cohen's distinctions between believing and accepting, or judging and premising, are illuminating in that they allow us to recognize the automatic—and non-agentive—character of thinking as a basic mental activity, while also allowing a form of reflexive, controlled thinking that leads a thinker to filter, select, reconsider (or deliberately misrepresent to others) her own judgments.

Tyler Burge and Christopher Peacocke have used a similar distinction to explore the conditions that make epistemic entitlement to self-knowledge possible. For such entitlement to be reached, one needs not merely to be in a position to reliably acquire beliefs about the world and oneself. One must in addition be able to critically appraise one's beliefs and change them in response to new reasons ('one must recognize reasons as reasons', Burge 1998: 246). A second requirement is that one should be able to self-ascribe one's thoughts

[22] Some philosophers seem to consider that the traditional expression 'mental act' automatically implies an agentive view of thinking. This assumption is however based on a misunderstanding of what 'act' (*actus*) means in the medieval, Thomist-Aristotelian sense of the term, where it is opposed to 'potentiality' (and not to 'passive reaction'). When Peter Geach (1957) adopted the concept of a *mental act* to refer to what psychological verbs such as 'see', 'hear', 'hope', 'think', are used to report, he did not mean that these verbs expressed mental *actions*. 'Act' as used by Geach is roughly synonymous with 'event', in opposition to 'power'. There is no action if there is no intention, trying, or volition whose semantic content is the goal to be achieved. In contrast with an action, an act in Geach's sense does not entail any purposeful involvement of the thinker in reaching the outcome of the 'act': the verb 'see' is a good example of a mental 'act' that does not involve a prior goal. The perceiver does not need to try to see to see.

[23] Cf. Cohen (1992: 12). Cf. also Engel (1998, 2000).

in a rational, non-contingent way. This second requirement can be spelled out in different ways. Burge favors the view that one must conceptually represent one's propositional attitudes as well as their contents. 'In critical practical reasoning, one must be able to—and sometimes actually—identify, distinguish, evaluate propositions conceptualized as expressing pro-attitudes, to distinguish them explicitly from those that express beliefs and to evaluate relations of reason among such propositions as so conceptualized' (pp. 247–8). In addition, being a critical reasoner crucially involves an ability to represent one's own self as a rational agent (p. 251). In this reflexive sense, *agency in thought is the capacity to revise one's own thoughts, as a permanent and immediate possibility,* in contrast with the simple notional and mediate possibility of influencing others' systems of beliefs (Burge 1998: 254–5). These preconditions for mental agency can also be seen as spelling out what motivates agents to perform mental actions. According to Burge, one does not passively endure the effects of reasons as one endures gravitational force. Reasons are rather (contentful) *motives* for attitudinal change: they authorize the thinker to either maintain or change her judgment. They furthermore motivate (*and not only cause*) the thinker to *immediately* shape or reshape her attitudes. As we have seen, such motivation is fueled by the reflexive recognition by a thinker of herself as aiming-at-truth. This view restricts mental action to subjects who are able to represent themselves as rational agents: non-human animals and younger children are automatically excluded.

Peacocke, on the other hand, alleviates the need for conceptual metarepresentation of first-order contents. He claims rather that, in self-ascribing mental states that involve the first-person essentially, such as actions, one can 'move rationally from a mental state to a self-ascription without representing oneself as enjoying that mental state.'[24] Peacocke however, as we saw, views concept possession as independent of personal-level conscious knowledge of the conditions for possessing the corresponding concepts (1999: 24 and 237). On his view, a thinker can be motivated to be sensitive to reasons even though he does not have the capacity of self-attribution of mental properties. As we shall see below (section 5), there are powerful reasons to favor a view that does not link rational motivation to conceptual self-attribution.

If believing and accepting have distinct epistemic roles, how can we characterize precisely why accepting is a manifestation of agency while believing is not? What precedes suggests an answer. Mental agency occurs in the context of critical reasoning. Isolated registerings do not qualify as expressions of agency when performed outside reasoning processes; they gain the status of active judgings

[24] Cf Peacocke (2007: section VI). In Peacocke (1999), the concept of *representational dependence* is introduced to account for cases in which a subject forms the belief 'I am F' by taking the associated mental state at face value. For example, the content of the mental state represents the subject as 'having some location in the spatiotemporal world' (p. 265).

when included in a reason-sensitive inferential process.[25] Let us summarize the agentive features of *judgment* as a critical activity.

1. A subject *tries,* successfully or not, to reach a rationally sound epistemic decision about a specific proposition: accept it as true, or reject it as false according to all the available relevant considerations that bear on the issue. Such an attempt entails a capacity to resist the pull to immediately register a fact, or uncritically jump from a set of premises to a conclusion.
2. Trying to judge both *constitutes a judging and causes an awareness of judging*: there is something it is *like* to judge, allowing the subject to apply to himself the concept of judgment, if it is available.
3. There are always, as in any action, *alternative courses one might have pursued*, that are generally not rational.[26] Choosing a strategy involves 'reference to rational standards'.[27]
4. Finally an individual mental act of judging may rationally motivate new actions: the thinker who has come to an epistemic decision about *P* may need, as a consequence, to confirm or modify her prior epistemic or conative commitments.

3. GENERALIZING TO MENTAL ACTIONS

How does this theory of agency in *reasoning* generalize to other forms of mental action? On a free reconstruction of Peacocke (2007), one could suggest that in every mental action:

1. A subject *tries,* successfully or not, to reach a psychological property subject to some pre-established norm (like truth, dependability, correctness, etc.) that he would not have reached otherwise.

[25] A similar view is developed in Peacocke (1998, 1999). An alternative view would maintain that judgings (subsuming an individual under a conceptual term, or stating that a concept is subordinated to another, or that a first-order concept belongs to the extension of a second-order concept) are mental operations that do not intrinsically possess any agentive feature. They are rather non-agentive building blocks in more extended processes that may qualify as actions.

[26] It has been objected that rational evaluation does not seem to involve any selection between alternatives: a thinker cannot and should not withhold a rational judgment that *P*; nor can she rationally decide what the content of her judgment will be. Therefore, judgings cannot be actions. As Peacocke correctly argues (1999: 19), this argument should be resisted. There are indeed many standard actions in which no rational choice is open to the agent: a driver does not choose how to drive safely—he chooses to drive safely. Given that thinking essentially involves an *inferential capacity*, telling which inferences are rational is not 'up to the thinker'.

[27] Burge 1998: 248. It requires, as Burge also insists, 'distinguishing subjectivities from more objectively supportable commitments' (1998: 248).

2. Trying to mentally A *constitutes* mentally A-ing and *causally contributes to* the awareness of doing so: there is something it is like to decide, calculate, and the like.
3. For any mental trying token, *alternative courses might have been pursued*. A given mental trying is subject to evaluation through rational standards.
4. Any individual mental act may in turn rationally motivate new mental actions. A thinker who has calculated that *P* may need, as a consequence, to revise the grounds for other acceptings, desirings, intentions, and plans.

We cannot elaborate on this definition for lack of space; we will rather concentrate on how it responds to the skeptical worry expressed in our introduction. We saw in section 1 that a definition of mental action that would exclusively rely on the discrimination by an agent of her being active or passive in a thinking episode would fail to be objectively applicable, for agents may be deceived about their own mental agency. Burge's view circumvents this difficulty, because his transcendental theory of mental agency makes the capacity to act mentally an a priori condition of rational agency rather than an empirical property. Mental agency is what makes a rational agent possible, and rational agents are posited as existing. Peacocke's view on mental action, however, is not transcendental. It aims at understanding how an individual subject can gain self-knowledge in mentally acting by gaining an *awareness of trying* to act mentally.[28] A second step consists in establishing the entitlement for a subject to make a rational transition, from his being aware of mentally A-ing, to judging that he is indeed doing so.

Thus, in order to understand how Peacocke responds to the sceptical worry above, we need first to understand how one is conscious of trying to A (to judge, decide, or imagine). Such awareness does not involve any perception nor proprioception, but rather 'the sense of agency from the inside'.[29] This experience has a crucial relation to the production of a corollary discharge.[30] On a common view, already sketched in section 1, a command being activated causes an efferent copy to be produced. Peacocke agrees: the agent indeed experiences agency for her subsequent (mental or bodily) action because of a corollary discharge signal being activated.[31] The very *existence* of a command, that is, of *a* trying, may suffice (if relevant error-feedback is prevented from

[28] If there is no *sense* of trying as a distinctive experience, in addition to the purely conceptual or functional features of trying, then indeed his definition of a mental action becomes circular: a mental action is or involves a trying; a trying, in turn, should only be defined as the common core of mental and bodily actions. To break this circle, trying has to be identified through an independent, subjective mode of access.

[29] Peacocke 2003 and 2007: s. II.

[30] i.e. an efferent copy of the command; that is, a neural signal that keeps track of a given command. See section 1 above.

[31] 'If the corollary discharge is caused by trying to perform the action in question, in normal subjects, that explains why, when there is no evidence to the contrary, trying itself causes an "apparent" action awareness' (in subjects with schizophrenia): Peacocke (2007).

being perceived) to produce apparent action-awareness. Other components of the experience include the sense of an (identification-free) first-personal and present-tensed mental action, made available through demonstrative reference.[32]

We can now appreciate how Peacocke deals with the sceptical worry articulated above. Action-awareness, on his view, whether in bodily or in mental action, 'should not be identified with any kind of belief, whether first- or second-order'.[33] Having the sense of acting is not *believing* that you are acting. A subject exposed to Wegner's experimental setting[34] can have a compelling feeling of agency, while also recognizing its illusory character. Having an awareness of trying, i.e. being conscious of acting rather than being acted upon, is a seeming, not a knowing. Obviously a belief can be formed on the basis of such seeming. But this belief is not presented by the theory as immune to error; one can be wrong when believing that one is engaging in a mental action (e.g. confuse one's imagining with one's judging, one's visualizing with one's remembering).[35]

Granting that one can have only apparent awareness that one tries to judge, say, when one actually unreflectively forms a belief, the second part of the anti-skeptical move is to explain why a subject is entitled to judging that she is mentally A-ing when she does, and, furthermore, that she is mentally A-ing that *P* rather than *Q*. Certainly, conditions 2, 3, and 4 above may explain why a thinker may reliably attribute to herself a mental action of a certain type and content: a subject who is first aware of *trying* to A subsequently becomes aware of *A-ing*, rather than Q-ing, has access to her reasons for A-ing, and is *motivated* to pursue further acts as a consequence of her A-ing. The set of causal-intentional relations that her trying maintains with her acting thus seems to allow reliable discrimination of genuine from illusory trying.

But how can entitlement to self-knowledge of mental action be secured on the basis of these relations? Generalizing from the epistemology of the perception to the domain of action seems to provide a response. Peacocke (2004) suggests that a perceiver having the experience as of *P* is entitled to judge that *P* on an a priori ground, which Peacocke calls the 'Complexity Reduction Principle'. It would be an a posteriori ground to merely claim that perception has the function of delivering reliable input to a belief system. A selectionist explanation justifies that, other things being equal, perceptual experiences are predominantly correct. And this justification becomes an entitlement if the selectionist explanation is itself warranted by the Complexity Reduction Principle (2004: chs. 3, 4). If a

[32] These components again reflect features of representation-dependent uses of self-attributive beliefs that can be, but as we will see, do not have to be, built on the basis of action-awareness. See n. 25.

[33] Peacocke 2007: s. I c.

[34] See Wegner (2003). The setting is contrived so as to lead a subject to have the impression that he did produce a given effect in the world, when he actually did not.

[35] One may also confuse this content of judging for that one. But this mistake does not threaten the very fact that a mental action was performed; it only allows such an action to be attempted and failed at, as it should be—all actions are subject to failure.

perceiver is entitled to believe that *P* when, other things being equal, she seems to see that *P*, the same seems to hold, *mutatis mutandis*, for an agent who has the apparent awareness of deciding, judging, or trying to memorize that *P*. She is entitled to believe herself to be A-ing—*ceteris paribus*—when it seems to her that she is A-ing.

Remaining Problems

The theory discussed has many interesting features, in particular the link that it establishes between mental action, self-awareness, and entitlement.[36] One of its main aims is to go beyond reliabilism by showing how mental action-awareness may, given the Complexity Reduction Principle, entitle a thinker to make true judgements about her own actions, and more generally, to perform rational mental actions (decidings, attendings, calculatings, etc.).

This principle, however, does not provide anything more than typically defeasible grounds for entitlement. The 'easiest explanation' of today usually becomes a false inference tomorrow; it is not clear how a substantive judgement about entitlement can result from a priori, very general considerations on how available explanations maximize simplicity. A second problem is to know how a non-sophisticated subject can appreciate the *relative ease* of the various alternative explanations for why I believe Φ when and only when I do.

Another difficulty for this account of mental actions is that it is not clear how a subject learns how to detect cases of self-deception with respect to whether she did perform a mental action. Such detection, as Peacocke observes, must be still more difficult in the case of a mental than a bodily action.[37] For example, a subject may believe wrongly that she succeeded in comparing two plans of action, when she in fact only considered one. She can nevertheless go on rehearsing apparent reasons to back up her decision, and develop new mental actions as a consequence. To rule out the permanent possibility of this kind of self-deception, we need to have clear markers, in the experience of mental agency, *of the way in which our token action satisfies the conditions of the type to which it belongs*. The general idea is that, for a mental action to be at all possible, the conditions for correction of a given mental action must be analogous to those prevailing in bodily action. If a subject did not have any principled entitlement to evaluate how a mental action is currently developing, or to retrodict how it was formed and what output it led to, no mental action could actually be performed rationally. We need to identify the basis on which the subject can judge (i) that she can perform A, (ii) that she has reached her goal or failed to reach it, or (iii) that she has performed no action.

[36] Another attractive feature is that it offers an account that holds for bodily and for mental actions. For a defense of this kind of account, see Proust (2006*a*).

[37] Peacocke 2007: s. II.

Finally, it is not clear, given the present definition of a mental action, what the range is of possible targets open to our mentally trying to achieve them. What is still in need of explanation is what distinguishes, in general, a type of mental action of the *directed* type (directed remembering, directed imagining, directed reasoning or computing) from mental operations of the automatic, *non-directed type* (automatic remembering, passive imagining, passive inference making). An adequate definition of a mental action should aim at articulating such a distinction, a basic requisite for any response to the sceptic's worry.

4. DIRECTED THINKING: A VOLITIONIST ACCOUNT

The structural differences between non-directed and directed thinking can be clarified by making explicit in non-phenomenological terms what willing, trying, or volition are.[38] Intuitively, willing an action A consists in trying to obtain typical effects, those effects being represented as reachable as a consequence of this willing.[39] This intuitive, subject-centered view of willing was already the target of the definitions provided in section 3. But it can be completed by an objective, process-centered definition: what distinguishes a non-willed movement from an action is that while the first is produced automatically, as a reaction to the environment, the second is selected as a function of context and individual preferences. What holds for bodily actions,[40] should hold for mental ones: automatic attending, registering, or deciding is a product of past conditioning, i.e. of associative memory. Active (or directed) attending, judging, or deciding consists in including the formerly automatic ability into a controlled sequence, thus using it as a means to a higher-level goal: some mental property. In directed memory, the higher-level goal is to retrieve a correct, specific memory; in directed decision, it is to come up with an adequate compromise between alternative goals and intentions; in directed attention, it is to allocate more processing resources to a first-order task.[41] To distinguish the directed mental event from the automatic,

[38] Three terms that I take to be equivalent: see Proust (2005: ch. 4). There are independent reasons that make such an analysis desirable. Proust (2005) argues that an analysis of action that distinguishes the capacity to execute an action from the capacity to represents one's reasons to act—as the volitionist approach does—is necessary to account for certain perturbations of action. There are well-known objections against a volitionist theory of action. They are discussed in detail in Proust (2001, 2005).

[39] See e.g. O'Shaughnessy (1980); Searle (1983); Peacocke (1998: 68). For a control view of action, see Proust (2005, 2006*a*) and Mossel (2005).

[40] The functional link between the two, according to a current view, is that when the automatic movement is stored in memory, i.e. represented in a 'motor lexicon', it can be used in new combinations in a means-to-goal sequence encompassing different contextual constraints. The former sequence becomes a lower order unit in a hierarchical control process. See Koechlin *et al.* (2003) and Koechlin and Jubault (2006). See also Shallice (1988).

[41] Theodule Ribot (1890) described such emergence of directed attention as 'Art constraining nature': where nature uses intense stimuli to automatically capture a subject's attention, 'art'

associative one, we can say that, in the former case, the operation is 'called', while in the second it is merely 'activated'.[42]

As a matter of definition, then,

> A *willing* or a *trying* is a mental event through which an operation from the repertory is (1) called because of its instrumental relationship to a goal, and (2) is thereby made available to executive processes.[43] In bodily action, the goal is ⌜that an external change be brought about in virtue of this trying⌝.[44] In mental action, the goal is ⌜that an epistemic—or motivational change be brought about in virtue of this trying⌝.

Let us observe the dual aspect of the content of willing in (1). On the one hand, one wills to achieve such and such a *distal goal* (e.g. my will is [that I remember A's last name in virtue of my willing]). This aspect is what drives the selection of a command of a certain type addressed to one's own system (Calculate! Plan! Remember!). On the other hand, one wills to achieve this token-reflexive goal *in a specific way*. This aspect constitutes the selection of a given pathway to the goal, i.e. 'how' to get there (e.g. I will remember by focusing my attention on the person whose name I am searching for). It is crucial, when appreciating the success of one's mental action (as well as of one's bodily action), to recognize that it was reached in the specific way that one's willing constitutively included.[45]

How is control causally efficacious? To answer this question, one needs to generalize the answer developed above for the case of motor action. In fact, I have shown elsewhere that this generalization is warranted by a priori functional considerations as well as by current neuroscientific evidence.[46] *Any* adaptive control needs some functional equivalent to forward models that help evaluate observed feedback by generating internal expectations.[47] Controlled

(human culture, pedagogy) recruits this ability to structure one's relation to the world according to one's own schedules and expectations.

[42] It may be tempting to interpret the call/activate distinction through the personal/subpersonal contrast. I don't think, however, that this contrast is really helpful, for it merely takes it for granted that the 'calling' process can only be triggered consciously, which there are serious reasons for doubting. Other reasons not to take this contrast at face value are offered below.

[43] This two-tiered structure of trying responds to the familiar puzzle, discussed by Ryle (1949), that if there is such thing as a mental act of heeding or trying, then there must of necessity be a second enslaved act: there is always something specific that you try to achieve, or something or other that you have to do to do it by way of attending. This however does not justify an adverbial theory of what a mental act is.

[44] A bodily action can also serve a mental goal: e.g. writing and reading are both bodily and mental capacities. We will see later that it is plausible that a mental action becomes conscious through bodily reafferences.

[45] A similar view was defended in Searle (1986) in the case of bodily action. On some difficulties of Searle's particular theory, see Proust (2003).

[46] Proust (2006*a*).

[47] As Roger Conant and W. Ross Ashby (1970) have shown, the most accurate and flexible way to control a system involves taking the system itself as a representational medium, i.e. simulating

thinking should similarly proceed by triggering self-simulations based on prior performance. In order to search one's memory in a controlled way (rather than by passively associating cues), one must be able to know whether one can reach cognitive adequacy in a reasonable length of time. A comparison must be performed between the known dynamics of successful retrieval and the present attempt at retrieval.[48]

Our definition above, if it is to help us understand the *source* of an individual's entitlement to self-knowledge, needs to be made explicit in two ways. Why is a subject able rationally to select a particular mental action and monitor it adequately? What are the *cognitive* conditions in which a mental action is 'called' and how do these conditions not only cause a mental action, but also guide a subject's *evaluation* of that action? A first condition has to do with the cognitive dynamics that make selection—'calling a command'—possible. You can only try to do something that you know how to do (in control theory terms: you can only select a command from your repertory). Such know-how is what allows you to direct yourself to attend, to judge, to plan, or to decide. This knowledge, however, is not theoretical; it rather manifests itself as a disposition to produce actions of that type. This disposition is constituted by having a repertory of forward models allowing for selection of that type of action. It is important to realize that developmental factors deeply condition our ability to develop mental actions, i.e. to do more than simply run automatic associative thoughts. Prior practice makes directed mental functioning of wider scope possible, by making a wider repertory available, and by motivating us to use it. What motivates us is that benefits can result from our mental action: this simply cannot be appreciated without prior experience of mental agency.

This can be made explicit as:

(1*a*) *Condition of control*: An agent will do P iff she knows how to produce P in a given motivational context. For example, an agent knows how to search for a name in memory only when developmentally mature enough to do so.

Our condition (1*a*) specifies condition (1). (1a), however, cannot be expressed properly without involving a 'motivational context', i.e. a set of desires and needs with the associated instrumental relations toward a goal. This context is structured by the present 'affordances' of the world as perceived, i.e. represented and categorized (e.g. I have been asked a question), or by such affordances simulated (remembered, imagined, inferred) in former thoughts (e.g. I associatively remember that I need to plan my next trip). The two other conditions for willing of action

the target action, using the system's own dynamics. In an optimal control system, the regulator's actions are 'merely the system's actions as seen through a specific mapping'.

[48] Perhaps the problem can be simplified by defining an 'envelope' of successful dynamics as viability theory suggests. I am indebted to Helena Frankowska for this point.

are 'executive conditions', in that they explain what makes know-how efficient in a context. They prominently involve motivational conditions:

(2*a*) *Condition of saliency:* A present motivational context makes *P* a *salient* goal for the agent. For example, a speaker predicts that the current conversation will lead her to refer to someone whose name presently escapes her.

Execution, however, consumes resources; there are many competing salient goals at any time, in most contexts; we therefore need to spell out a second executive condition, namely:

(2*b*) *Condition of quantity of motivation*: Motivation must be *sufficient* to allow the agent to produce *P* in a controlled way. For example, you may be too tired or too hurried to search your memory for a proper name.

Our completed definition should now provide us with an answer to the two questions raised above—questions whose answers will allow us to later explain the link between mental action, awareness, and entitlement. A subject can select and monitor a particular mental action because it belongs to her control repertory. When however, a mental action from an individual's repertory is called, it can by the same token be simulated, evaluated, or predicted for adequacy. Part of this activity is performed unconsciously, just as the preparation of a bodily action (which also involves simulation and evaluation) is shown to be performed outside awareness. It is hypothesized by scientists, rather than experienced by subjects, that a set of *comparators* allows one to *anticipate* how things normally develop for such and such a type of mental action (say: a directed remembering, or a planning). These anticipations are then compared with actual feedback. For example, comparing the dynamics of my present attempt at retrieving a proper name to a baseline allows me to predict that I can remember that name within a few seconds.

This last sentence expresses the dual aspect of control and monitoring. On the one hand, control is exercised by our brains rather than by our conscious selves. On the other, there is something it is like for us to perform a given mental action. One can feel that one is about to succeed at it, or is bound to fail. One can feel that one is deeply engaged in the action or only marginally interested in getting things right. One might further consciously reason about the consequences of attaining versus failing the goal of this action, given these anticipations. The issue of mental action selection ('why should I attend, judge, plan, remember?') cannot thus be properly analyzed without also taking into consideration the subjective monitoring that is part of any mental trying. Is this subjective monitoring always consciously performed? This is an issue that is not answerable at this stage of consciousness studies. If it turns out that there are several degrees in conscious awareness, with an intermediate form that consists of having subjective phenomenological states with no concomitant attention, and/or no capacity to report on it, we will have to accept that an agent may have

a form of implicit, but still phenomenological, access to how an action feels, without this fact preventing adequate control from occurring.

Under what conditions, then, is a thinking episode under the *agent's* control? Does the answer lie in the awareness the agent has *that* she is in control of her mental action, in particular in her awareness of being active rather than passive with respect to a given thought? Some philosophers would certainly insist that the *agent's* being in control presupposes that she exercises her control *in a conscious way*. On this view, only a thinker who is conscious of her ability to control her own thinking would qualify as a mental agent. One of the problems of this view is that it supposes that there is a clear delimitation between the personal and subpersonal levels, which is not established; indeed a nominal distinction between 'personal and subpersonal levels' certainly does not suffice to prove that there is clear demarcation.[49] Although it has been found intuitive, it is not clear that it is functionally robust, or stable. We may become conscious of having performed a bodily action after it was performed (having ducked to avoid a projectile). An agent may also only have awareness prior to action rather than when the action occurs, particularly in the case of mental actions. Suppose I have consciously planned at t1 to remember A-ing at t2: I am now at t2, I now remember A-ing but have no simultaneous awareness of having planned to do so.[50] So I may have acted mentally without having the current ability to recognize that I was indeed acting rather than associating. To be clear about what is required for the agent to control her mental actions, we need to turn back to the condition of control above. Having procedural, even implicit, reflexive knowledge of what one can do is certainly the basic cognitive condition for an ability to control one's thoughts. But such procedural knowledge need not be displayed in any other way than calling one relevant action from the repertory (as articulated in the condition of control above), especially if we consider mental agents with no conceptually informed self-awareness.

The same seems to be true for controlled thinking performed by human agents. The case for subpersonal, largely unconscious, mechanisms that play a role in rational evaluation can be made by showing how *feelings* and *affective states* can orient our thinking in crucial ways. The point has been convincingly expressed by Christopher Hookway (2003: 79), that 'we can be confident of the rationality of our beliefs only if we can be confident of our habits of ''immediate evaluation''.' Such confidence is a conscious feeling generated by unconscious, subpersonal processes; it is immediately available to the subject; it has perfectly definite evaluative content. The domain of directed thinking offers us many such examples of a 'cross-over' mode of operation, through which subpersonal processes correlate with personal-level, conscious markers.[51] I will now attempt to show that the agent's having access to such cross-over markers constitutes

[49] See Pacherie and Proust (2008). [50] See Golwitzer (2006).
[51] See Koriat's (2000) analysis of 'noetic' feelings.

a crucial dimension of awareness of agency. Entitlement to self-knowledge for mental agency essentially depends on it.

5. METACOGNITIVE FEELINGS AND ENTITLEMENT TO MENTAL AGENCY SELF-ATTRIBUTION

Metacognitive Feelings

Our first task is to examine what metacognitive feelings are, and how they relate to self-knowledge. As we saw in the preceding section, comparators deliver evaluations either of the *anticipated* need and feasibility of our mental actions, or of the *observed* results attained by performing them. Now, the mental agent has immediate *subjective, phenomenological* access to the comparator's verdict. Such direct access is made possible through dedicated feelings.[52] All of them express the degree of subjective epistemic uncertainty or sensed feasibility for a given task or outcome.[53] For this reason, these feelings are also called 'metacognitive' or 'epistemic'. The value of all these feelings is 'epistemic' because their function is to help a thinker recognize and evaluate the dynamics of her own beliefs, memories, and plans, with respect to truth or adequacy. These feelings are thus endowed with representational content, as is arguably the general case for emotions. As we shall see below, they represent the cognitive adequacy/inadequacy of a specific mental action. They don't need to involve, however, a conceptual representation of one's having beliefs or other mental states, nor of their truth and falsity.[54] But, as feelings, they also have a motivational dimension, which does involve a capacity to regulate one's affects through proper commands.

A few examples will help us to recognize the wide scope of this form of awareness of mental action. In a 'tip of the tongue' experience, a subject becomes aware both that she is failing to retrieve a memory, and that it is worth trying harder. Feelings of knowing, or not knowing, or vaguely knowing, are other forms of experience that arise when one evaluates directed learning.

[52] Hookway (2003) convincingly argues that epistemic evaluation involves immediate phenomenological states.

[53] Some express 'velocity' aspects of the cognitive dynamics (the sense of being able to achieve a cognitive goal). Some specifically express 'acceleration' aspects of the predicted mental dynamics (the sense of insight and/or of exhilaration in reaching an important finding are good examples, in contrast with the feeling of 'sinking' when one finds oneself cognitively helpless in an important matter). On this difference, see Carver and Scheier (1998).

[54] Having epistemic feelings allows a subject to immediately form an evaluation or a prediction about her past or future mental action. It appears that such feelings can be entertained in the absence of a metarepresentation to the effect that one remembers, judges, decides that *P* (Proust 2007). This suggests that awareness of one's own mental action is not *primarily descriptive*, as in ⌜I now judge/remember that *P*⌝. It may rather primarily involve a normative assessment of what we conceptually analyze as ⌜I am confident/unsure that my judgment/memory that *P* is correct⌝.

There are many more types,[55] associated with reasoning, planning, deciding, such as regret for having made a decision, a 'rational' feeling now studied in neuroeconomics.[56]

Metacognitive feelings, in their variety, seem to track the norms that constrain the efficiency of the thinking, information processing system. A range of mental acts has to do with perceptual intake and subsequent perceptual belief fixation. The associated metacognitive feelings primarily track 'informational quality', i.e. the optimal signal-to-noise ratio of a sensorimotor, perceptual, or recreative imagination operation.[57] More generally, mental actions must be deployed in areas such as judgment, reasoning, and decision. In this general case, the metacognitive feelings track 'cognitive adequacy', i.e. the correct evaluation of the resources available/needed for a given mental task, of such and such import.[58] (As shown by researchers in metamemory, a human agent with no specialized training[59] can learn how to set her own decision thresholds in the most rational way.)

Given the superposition, in adult humans, of a metacognitive and a metarepresentational capacity,[60] it may seem difficult to tease apart the role of immediate epistemic feelings and of conceptual recognition by a thinker that she is aiming at truth, as constituting the motivating force that drives mental actions. Let us note, however, that even though motivation to reassess one's judgments may originate in such an explicitly reflexive way, it does not have to. Empirical evidence suggests

[55] Hookway lists many cases of affective states that regulate reasoning. Note that the central notion of relevance needs mental effort to be assessed and compared, which plausibly also involves a metacognitive feeling. See Proust (2008).

[56] See Camille *et al.* (2004).

[57] Does our bipartition exhaust the varieties of mental action? One might object that it fails to cover a major kind of mental action, which Peacocke calls 'directed thought': in this form of thinking, a subject has 'the intention to think a thought which stands in certain relations to other thoughts' (Peacocke 1998: 67). In preparing a response, for example, a philosopher needs to collect relevant ideas, and organize them rationally and persuasively. This kind of case does not, however, constitute a new category of mental action. It does involve planning of successive episodes of directed memory search, evaluation, and pruning of irrelevant topics. All the elementary mental actions involved belong to the two categories already described. Some of them involve increasing the informational quality of thoughts by filtering out those thoughts that are foreign to the task, while allowing the relevant ideas to be 'in full view' (by writing or rehearsing them, etc.); the rest aim at increasing the cognitive adequacy of one's reasoning by selecting, among competing solutions, the one that best satisfies a set of constraints given the task at hand. What is not up to the thinker, again, is which ideas are available for selection at a given effort level, and which one to choose among alternative candidates. Thus, again, it would not be correct to describe a mental action as that of picking 'this' thought, because the thought to be retained and worked out is not deliberately judged prominent or relevant. But our discussion shows that, under some understanding of that description, it is correct to say with Peacocke that the subject intends to 'think a thought with certain relations to other thoughts': these relations however are not meaning relations, but rather 'adequacy relations', such as: picking something that will turn out to be convincingly relevant to the task, and picking it in a way that allows that relevance to be itself correctly evaluated.

[58] There is thus an overlap between informational quality and cognitive adequacy, as illustrated in signal detection theory.

[59] As well as a few other mammals (see Smith *et al.* 2003).

[60] See Proust (2007).

that metacognition occurs in nonhuman animals (such as macaque monkeys) which seem to form no self-representations nor to have metarepresentational access to epistemic norms. Although this remains a controversial issue, metacognitive capacities may also appear in human children before they have a theory of mind, i.e. a capacity to represent their own mental states and those of others.

Let us sum up. Epistemic or metacognitive feelings express the cognitive adequacy of an *anticipated or executed* mental action. In both cases, a subject represents, in a qualitative, non-conceptual way, the reliability of the information on which her mental action depends in each of its relevant parameters (e.g. the vivacity of her memory, the precision of her perception, the strength of her motivation). These feelings contribute to guiding an upcoming mental action (possibly leading one to think what these words express: 'I should stop trying to remember, this name is not in my memory'). They also contribute to assessing, and possibly replanning, a mental action already performed (e.g. 'my judgment/decision does not feel right').

Epistemic Feelings and Entitlement

We can now return to the issue of entitlement to first-person knowledge of mental agency. As we saw in section 3, the experience of acting may be a source of self-knowledge through apparent action-awareness. But we found that it was unclear whether such prima facie awareness could entitle a subject to believe that she is *in fact* acting mentally: laymen and philosophers have discordant views on this issue; patients with schizophrenia tend to attribute agency for their own thoughts to others. Entitlement provides a form of externalist justification, which itself requires that there is no objective uncertainty about the ground of the entitlement. Typically, although the subject has no explicit reason available to back up her belief that *P*, she has an experience that immediately compels her to form the associated belief. She furthermore has no explicit reason *not to* trust this way of forming a belief. The domain of mental action may seem to resist this strategy for two reasons. First, a subject who acts mentally seems by definition to have control over what she does: in contrast with what happens when a subject merely perceives what is the case, a subject who, say, attends to her perceiving that *P* has a way of establishing internal criteria for having, or not having, perceived correctly. Second, as we just noted, there does not seem to be any regularly compelling experience of agency while engaging in directed thinking.

The response to these two objections is that the objector has conflated two different sources of awareness for a mental action. When you try to remember that *P*, you control an operation of remembering that might, otherwise, not be automatically activated. What you control is not: the outcome of the operation of remembering (say: the fact that Jane's daughter is called 'Mary'), but the disposition to retrieve an item from memory. So when you retrieve a name from memory, you are presented with a fact, ⌜Jane's daughter is called Mary⌝,

and you're entitled to say that you remember that name. This is not, however, the proper level for an awareness of *agency*. For you could have been presented with the same fact, and be similarly entitled to say that you remember Jane's daughter's name, even if your memory had been prompted automatically, without any control or trying. The proper level at which you feel agentive is when you assess your *capacity to act (predictively or retrospectively).* Epistemic feelings, then, present you with facts (I know/I don't know this name), that are essential *motivators* for triggering relevant mental actions and essential *criteria for appraising* their outcomes. Thus, when having epistemic feelings, one is typically in a situation analogous to the first-order case of remembering or perceiving as far as entitlement goes. But the conceptual content of the entitlement is the judgment ⌜that I am presently evaluating my (past or future) A-ing⌝.[61] In metacognitive feelings, you are given a non-conceptual, 'procedurally reflexive' equivalent of that conceptual metarepresentation of entitlement.

6. CONCLUSION

Let me summarize the conclusion to be drawn from our discussion. On the view presented here, a subject is entitled to a sense of agency for a particular mental action if the metacognitive feedback relevant for that action both causes and justifies her performing that action (or, in the retrospective case, if it allows her to evaluate the success of her action). On this view, the ground of one's entitlement to self-knowledge about one's mental actions consists in dynamic facts that allow metacognitive feedback to be generated and used flexibly as a rational norm for how to perform a kind of mental action. Action awareness is constituted by an exquisite, dynamically informed sensitivity to one's own cognitive adequacy as manifested in a mental action token.

We saw in the introduction to this chapter that the issue of thought insertion and the variability of intuitions in normal subjects were prominent reasons not to take awareness of agency as a prima facie entitling condition for believing that one acts mentally. The question naturally arises of how the present definition of a mental action accounts for perturbations and variations in the sense of agency. Given a control view of the mental states involved, such perturbation or variation might be the case either (i) if a subject suddenly proves unable to control her own thoughts, or (ii) if she does not get the appropriate feedback when she does, or (iii) if she does not appropriately use the feedback she receives while monitoring a directed thinking event. What is known from schizophrenic delusions suggests that (iii) may be the correct hypothesis. Let us briefly consider why.

[61] It would be worth examining how the dynamics of prior mental actions of the same type is a calibrating precondition for enjoying occurrent feelings bearing on new tokens of thinking. Space limitation does not allow us to pursue this reflection here. See Proust (2009).

It is known that control (namely, the ability to call a specific command to regulate one's thinking processes) is somewhat altered in patients with schizophrenia, but apparently not to the point of fully disorganizing the patients' thinking processes, as may be the case in dementiae. The second hypothesis (ii) is not to be excluded, given the well-known 'flattening of affect' that is part of the schizophrenic syndrome. If, indeed, metacognitive feedback is conveyed by affective states, then one might speculate that patients have trouble producing or reading emotional signals serving metacognition as they do for those serving social needs. But the third hypothesis (iii) is also quite plausible at this point. It has been shown by researchers in metamemory that patients with schizophrenia tend to have a reduced 'control sensitivity': they fail to regulate their commands on the basis of the metacognitive feelings that they do feel (Bacon *et al.* 2007). An account like the one presented above is compatible with these findings. If metacognitive feelings are a necessary structure for feeling responsible for one's mental actions, then an inability either to have them in the normal way, or to use them in controlling rememberings, plannings, decidings, etc., should deeply disturb the sense of agency for thoughts as well as for bodily actions.

Our explanation also accounts for the variations in sense of agency for thought in normal subjects. On the view defended here, epistemic feelings are *given* to a thinker rather than controlled: feelings of knowing 'just happen'. Furthermore, they are given in a *non-conceptual* format: they have a specific motivational content that can be redescribed through concepts. Such redescription however does not seem to be necessary for the epistemic feelings to fulfill their function. Therefore, there is scope for variation in the amount of redescription of their thinking processes that subjects need to perform. The variation is inter alia modulated by the amount of inferential self-knowledge they expect to gain on the basis of their metacognitive evaluation.

The claim made here, that metacognitive thinking is a main provider of self-knowledge 'as a thinker' (rather than 'as a physical agent' or 'as a social agent'), is in agreement with current views[62] that self-knowledge can be seen, in central cases, as a product of rational agency. It does not follow from this claim, however, that having a self is a precondition for implicitly categorizing attitudes. A self *develops*, rather, from a capacity to implicitly revise one's attitudes through appropriate metacognitive processes.[63] Granted that tokens of automatic judging do not involve epistemic agency, a subject becomes implicitly aware that she judges rather than imagines, say, when the question of the epistemic adequacy of each process is raised. A sence of epistemic agency appears only when regulation becomes decision-sensitive.[64]

[62] See in particular Moran (2001) and O'Brien (2007). [63] See Proust (2003*b*).

[64] I wish to express my gratitude to Lucy O'Brien, Matthew Soteriou, Dick Carter, and two anonymous reviewers for their comments on a previous version. I have an additional debt to Dick for his linguistic revision. I am also indebted to Jérôme Dokic, Elisabeth Pacherie, and the members

REFERENCES

Bacon, E., M. Izaute, and J. M. Danion (2007) 'Preserved Memory Monitoring But Impaired Memory Control during Episodic Encoding in Patients with Schizophrenia', *Journal of the International Neuropsychological Society,* 13: 219–27.

Blakemore, S. J., and J. Decety (2001) 'From the Perception of Action to the Understanding of Intention', *Nature Reviews Neuroscience* (Aug.): 561–7.

Boyer, P. (2002) *Religion Explained: The Evolutionary Origins of Religious Thought* (New York: Basic Books).

Burge, T. (1998*a*) 'Reason and the First Person', in C. Wright, B. C. Smithy, and C. Macdonald, *Knowing Our Own Minds* (Oxford: Oxford University Press), 243–70.

Burge, T. (1998*b*) 'Individualism and the Mental', in P. Ludlow and N. Martin (eds.), *Externalism and Self-Knowledge* (Stanford, CA: CSLI Publications), 21–83.

Camille, N., G. Coricelli, J. Sallet, P. Pradat-Diehl, J. R. Duhamel, and A. Sirigu (2004) 'The Involvement of the Orbitofrontal Cortex in the Experience of Regret', *Science,* 304/5674 (21 May): 1167–70.

Campbell, J. (1998) 'Le Modèle de la schizophrénie de Christopher Frith', in H. Grivois and J. Proust (eds.), *Subjectivité et conscience d'agir: Approches cognitive et clinique de la psychose* (Paris: Presses Universitaires de France), 99–113.

____ (1999) 'Schizophrenia, the Space of Reasons, and Thinking as a Motor Process', *The Monist*, 82/4: 609–25.

____ (2001) 'Rationality, Meaning and the Analysis of Delusion', *Philosophy, Psychiatry and Psychology*, 8: 89–100.

____ (2002) 'The Ownership of Thoughts', *Philosophy, Psychiatry and Psychology*, 9/1: 35–9.

Carver, S. C., and M. F. Scheier (1998) *On the Self-Regulation of Behavior* (Cambridge: Cambridge University Press).

Cohen, J. (1992) *An Essay on Belief and Acceptance* (Oxford: Oxford University Press).

Conant, R. C., and W. R. Ashby (1970) 'Every Good Regulator of a System Must Be a Model of That System', *International Journal of Systems Science*, 1: 89–97.

Damasio, A. R., B. J. Everitt, and D. Bishop (1996) 'The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]', *Philosophical Transactions; Biological Sciences*, 351/1346: 1413–20.

Daprati, E., N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Dalery, and M. Jeannerod (1997) 'Looking for the Agent: An Investigation into Self-Consciousness and Consciousness of the Action in Patients with Schizophrenia', *Cognition,* 65: 71–86.

Davidson, D. (1980) *Essays on Actions and Events* (Oxford: Oxford University Press).

Dawkins, R. (1976) *The Selfish Gene* (Oxford: Oxford University Press).

Desimone, R. and J. Duncan (1995) 'Neural Mechanisms of Selective Visual Attention', *Annual Review of Neuroscience*, 18: 193–222.

Dretske, F. (2000) 'Entitlement: Epistemic Rights without Epistemic Duties?', *Philosophy and Epistemological Research*, 60/3: 591–606.

Engel, P. (1998) 'Believing, Holding True, and Accepting', *Philosophical Explorations*, 1/2: 140–51.

of the APIC seminar, to Pierre Livet, to Achim Stefan and his group in Osnabrück, and to the SOPHA 2006 audience for very helpful discussions.

Engel, P. (ed.) (2000) *Believing and Accepting* (Dordrecht: Kluwer).
Farrer, C., and C. D. Frith (2002) 'Experiencing Oneself vs. Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency', *NeuroImage*, 15: 596–603.
Farrer, C., N. Franck, N. Georgieff, C. D. Frith, J. Decety, and M. Jeannerod (2003) 'Modulating the Experience of Agency: A Positron Emission Tomography Study', *NeuroImage*, 18: 324–33.
Feinberg, I. (1978) 'Efference Copy and Corollary Discharge: Implications for Thinking and its Disorders', *Schizophrenia Bulletin,* 4: 636–40.
Frankfurt, H. (1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).
Frith, C. D., S.-J. Blakemore, and D. M. Wolpert (2000) 'Explaining the Symptoms of Schizophrenia: Abnormalities in the Awareness of Action', *Brain Research Reviews*, 31: 357–63.
Gallagher, S. (2000) 'Self-Reference and Schizophrenia', in D. Zahavi (ed.), *Exploring the Self* (Amsterdam: John Benjamins), 203–39.
Geach, P. (1959) *Mental Acts: Their Content and their Objects* (London: Routledge & Kegan Paul).
Gerrans, P. (2001) 'Authorship and Ownership of Thoughts', *Philosophy Psychiatry and Psychology*, 8/2–3: 231–7.
Goldman, A. (1993) 'The Psychology of Folk Psychology', *Behavioral and Brain Sciences,* 16: 15–28.
Gollwitzer, P. M., and P. Sheeran (2006) 'Advances in Experimental Social Psychology, Implementation Intentions and Goal Achievement: A Meta-Analysis of Efects and Processes', in *Advances in Experimental Social Psychology* (London: Academic Press), 69–120.
Graham, G. (2007) 'Thinking Inserted Thoughts', in J. Radden (ed.), *Oxford Companion to the Philosophy of Psychiatry* (Oxford and New York: Oxford University Press), 89–105.
Hookway, C. (2003) 'Affective States and Epistemic Immediacy', *Metaphilosophy*, 34: 78–96; repr. in Michael Brady and Duncan Pritchard (eds.), *Moral and Epistemic Virtues* (Oxford: Blackwell), 75–92.
Hume, D. (1739/1962) *A Treatise on Human Nature*, ed. A. Selby-Bigge (Oxford: Oxford University Press).
Jackson, J. H. (1958) *Selective Writings* (New York: Basic Books).
Knoblich, G., M. Öllinger, and M. Spivey (2005) 'Tracking the Eyes to Obtain Insight into Insight Problem Solving', in G. Underwood (ed.), *Cognitive Processes in Eye Guidance* (Oxford: Oxford University Press), 355–75.
Koechlin, E., and T. Jubault (2006) 'Broca's Area and the Hierarchical Organization of Human Behavior', *Neuron*, 50/6: 963–74.
____ C. Ody, and F. Kouneiher (2003) 'The Architecture of Cognitive Control in the Human Prefrontal Cortex', *Science*, 302 (14 Nov.): 1181–5.
Koriat, A. (2000) 'The Feeling of Knowing: Some Metatheoretical Implications for Consciousness and Control', *Consciousness and Cognition*, 9: 149–71.
Lormand, E. (1996) 'Nonphenomenal Consciousness', *Noûs*, 30/2: 242–61.
MacKay, D. M. (1966) 'Cerebral Organization and the Conscious Control of Action', in J. C. Eccles (ed.), *Brain and Conscious Experience* (New York: Springer Verlag), 422–45.

Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press).

Mossel, B. (2005) 'Action, Control and Sensations of Acting', *Philosophical Studies*, 124/2: 129–80.

O'Brien, L. (2007) 'Self-Knowledge, Agency and Force', *Philosophy and Phenomenological Research*, 71/3: 580–601.

O'Shaughnessy, B. (1980) *The Will* (2 vols. Cambridge: Cambridge University Press).

Pacherie, E., and J. Proust (2008) 'Neurosciences et compréhension d'autrui', in E. Ennen, L. Faucher, P. Poirier, and É. Racine (eds.), *Des neurones à la philosophie: Neurophilosophie et philosophie des neurosciences* (Paris: DeBoeck Université), 295–328.

Peacocke, C. (1986) *Thoughts: An Essay on Content* (Oxford: Basil Blackwell).

____ (1998) 'Conscious Attitudes, Attention and Self-Knowledge', in C. Wright, B. C. Smithy, and C. Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 63–98.

____ (1999) *Being Known* (Oxford: Clarendon Press).

____ (2003) 'Awareness, Ownership, and Knowledge', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness: Issues in Philosophy and Psychology* (Oxford: Oxford University Press), 94–110.

Penfield, W. (1974) 'The Mind and the Highest Brain-Mechanism', *American Scholar*, 43: 237–46.

Proust, J. (2000) 'Awareness of Agency: Three Levels of Analysis', in T. Metzinger (ed.), *The Neural Correlates of Consciousness* (Cambridge, MA: MIT Press), 307–24.

____ (2001) 'A Plea for Mental Acts', *Synthese*, 129: 105–28.

____ (2003*a*) 'Perceiving Intentions', in J. Roessler and N. Eilan (eds.), *Agency and Self-Awareness: Issues in Philosophy and Psychology* (Oxford: Oxford University Press), 296–320.

____ (2003*b*) 'Thinking of Oneself as the Same', *Consciousness and Cognition*, 12/4: 495–509.

____ (2003*c*) 'Action', in B. Smith (ed.), *John Searle* (Cambridge, MA: Cambridge University Press), 102–27.

____ (2005) *La Nature de la volonté* (Paris: Folio-Gallimard).

____ (2006*a*) 'Agency in Schizophrenics from a Control Theory Viewpoint', in W. Prinz and N. Sebanz (eds.), *Disorders of Volition* (Cambridge, MA: MIT Press), 87–118.

____ (2006*b*) 'Rationality and Metacognition in Non-Human Animals', in S. Hurley and M. Nudds (eds.), *Rational Animals*? (Oxford: Oxford University Press), 247–74.

____ (2007) 'Metacognition and Metarepresentation: Is a Self-Directed Theory of Mind a Precondition for Metacognition?', *Synthese*, 2: 271–95.

____ (2008) 'Metacognition in Conversation', in I. Wachmuth and G. Knoblich (eds.), *Embodied Communication* (Oxford: Oxford University Press), 329–56.

____ 2009 'Epistemic Agency and Metacognition: An Externalist View', *Proceedings of the Aristotelian Society*, CVIII, 3, 291–268.

Recanati, F. (2000) *Oratio Obliqua, Oratio Recta* (Oxford: Blackwell).

Ribot, T. (1890) *The Psychology of Attention* (Chicago, IL: Open Court Publishing Co.).

Ryle, G. (1949) *The Concept of Mind* (London: Hutchinson).

Searle, J. R. (1983) *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press).

Shallice, T. (1988) *From Neuropsychology to Mental Structure* (Cambridge: Cambridge University Press).

Smith, J. D., W. E. Shields, and D. A. Washburn (2003) 'The Comparative Psychology of Uncertainty Monitoring and Metacognition', *Behavioral and Brain Sciences*, 26/3: 317–73.

Sosa, E. (1991) *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press).

Spence, S. A. (2001) 'Alien Control: From Phenomenology to Cognitive Neurobiology', *Philosophy, Psychiatry, and Psychology,* 8/2–3: 163–72.

Sperber, D. (1996) *Explaining Culture: A Naturalistic Approach* (Oxford: Blackwell).

Sperry, R. W. (1950) 'Neural Basis of the Spontaneous Optokinetic Response Produced by Visual Inversion', *Journal of Comparative and Physiological Psychology,* 43: 482–9.

Sterelny, K. (2003) *Thought in a Hostile World: The Evolution of Human Cognition* (Oxford: Blackwell).

Strawson, G. (2003) 'Mental Ballistics or the Involuntariness of Spontaneity', *Proceedings of the Aristotelian Society*, 77: 227–56.

Velleman, J. D. (2000) *The Possibility of Practical Reason* (Oxford: Clarendon Press).

Von Holst, E., and H. Mittelstaedt (1950) 'Das reafferenzprinzip', *Naturwissenschaften*, 37: 464–76.

Vygostky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes* (Cambridge, MA: Harvard University Press).

Wegner, D. M. (1994) 'Ironic Processes of Mental Control', *Psychological Review*, 101: 34–52.

Williams, B. (1971) 'Deciding to Believe', in *Problems of the Self* (Cambridge: Cambridge University Press), 136–51.

Wolpert, D. M., Z. Ghahramani, and J. R. Flanagan (2001) 'Perspectives and Problems in Motor Learning', *Trends in Cognitive Sciences,* 5/11: 487–94.

This page intentionally left blank

Index