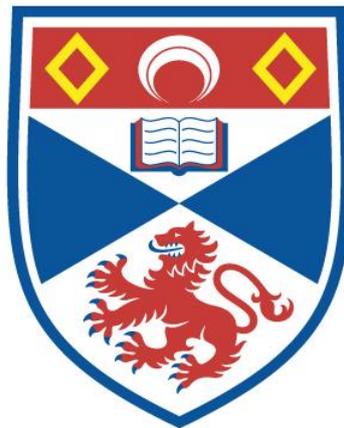


THE UNITY OF ACTION

Janice Chik

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2014

Full metadata for this item is available in
Research@StAndrews:FullText
at:

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/6550>

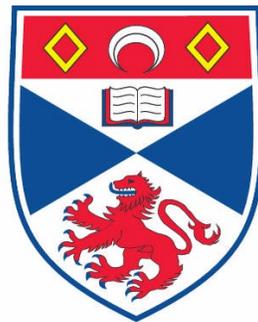
This item is protected by original copyright

This item is licensed under a
Creative Commons Licence

The Unity of Action

Janice Chik

A thesis submitted for the degree of PhD
at the
University of St Andrews



2014

THE UNITY OF ACTION

Janice Chik

Thesis for the degree of
Doctor of Philosophy (Philosophy)
University of St Andrews, UK
September 2014

Supervisors:

Professor Sarah Broadie (Philosophy)

Professor John Haldane (Philosophy)

Abstract

This thesis develops a disjunctivist approach to action as an alternative to the standard causal theory, or 'causalism'. The standard theory promotes a concept of action as constituted by a bodily event joined to certain mental conditions by a bond of causation. A disjunctivist approach, in contrast, claims that action must be distinguished by more than merely its etiology: action and mere movement are fundamentally different kinds. Recent objections to the causal theory of action are first surveyed, and the common causalist assumption claiming Aristotle as the progenitor of the causal theory is examined and dismissed. More refined interpretations of Aristotle's thought on action yield two different concepts: action as change, and action as a unified psychophysical process. The latter in particular is argued to hold promise as a basis for developing the disjunctivist approach to action. The remainder of the thesis therefore considers a contemporary account of psychophysicality, known as 'embodiment theory' (Hanna and Maiese 2009), with the conclusion that the intelligibility of the account depends on appeal to a recent variant of top-down causation (Steward 2012). The thesis also concludes that consideration of the concept of an animal agent makes it entirely unsurprising that the mental and physical are always found together in voluntary movement, and that the embodiment theory's central notion of 'property fusion' potentially complements a naturalistic variant of top-down causation in explanations of agency.

Acknowledgements

For generous financial support of my PhD research conducted from 2010-2013, I wish to firstly thank the UK Government (particularly the Scottish Funding Council) for the Scottish Overseas Research Student Award, as well as the University of St Andrews, for tuition and maintenance scholarships. I am also grateful to the Royal Institute of Philosophy for awarding me a research bursary for 2013-2014. For 2009-2010, in which I spent a year of independent research in Rome prior to commencing studies at the University of St Andrews, I owe sincere thanks to Luis Tellez at the Witherspoon Institute in Princeton, and to the Santa Croce Foundation.

At the University of St Andrews, I have benefited from conversations on early drafts with Jens Timmermann, Katherine Hawley, and Simon Prosser. I am particularly grateful to John Haldane, my secondary supervisor, for many long exchanges and for his helpful comments and advice on drafts. I wish also to thank Adrian Haddock (Stirling/St Andrews) and Helen Steward (Leeds), for their thoughtful questions and suggestions during the *viva voce*. My greatest appreciation in this regard is owed, of course, to Sarah Broadie, my primary supervisor. My debt to her is immense, not only for her patient and meticulous guidance of this thesis, but also for instructing me in certain philosophical habits; and most of all, for instilling within me the lifelong determination to make sustained progress in perfecting such habits, as one would in the virtues necessary for any worthy craft.

I should also record my gratitude to the following individuals who have played a crucial role in my intellectual formation outside of St Andrews: Al Martinich, Alex Mourelatos, J Budziszewski, and Dan Bonevac, at the University of Texas at Austin; as well as David Solomon and John O'Callaghan at the University of Notre Dame. In particular, I am greatly indebted to Rob Koons (UT-Austin), for his ongoing mentorship, friendship, and professional support; and to Fr Stephen Brock (Santa

Croce, Rome), for preparing me rigorously for PhD research at St Andrews, and for teaching me, *inter alia*, a deep and lasting appreciation for reading classical texts.

This work would not have been possible without the faithful friendship and frequently philosophical camaraderie of many. In St Andrews, I remain indebted to Lorn and Mary Macintyre, Fr Andrew Kingham, and various friends from Deans Court: especially Felix Pinkert, Hauna Ondrey, William Simpson, and Jess Volz. I am also most grateful to friends made in Glasgow, especially Lesley Hickson, Jane Mycock, and the ladies at Glenalvon. In Cambridge (UK), where I spent a year of research, I recall with gratitude my friendships there: with Amal Marogy, Fiona Doherty, Mark and Lucia Retter, Nicholas Teh, Carlo Rossi, and the action theory reading group with Jane Heal. My year researching in Princeton was similarly productive thanks to the supportive fellowship of many, including Corazon and Anselm Jeevaratnam, Gabby and Sherif Girgis, Cassy and Patrick Hough, Jennifer Bryson, Rosalie Stoner, and the ladies of Wiggins House, especially Barbara Falk. I also owe thanks to the Witherspoon Institute for allowing me to use the library and seminar room at Whelan Hall for writing. In Rome, I remain affectionately appreciative of *le donne di Celimontano*, especially Emanuela Genovese, Silvia DiPastena, Marta Rocchi, and Monica Pendolino; and with special gratitude, Fr Giulio Maspero.

My parents, Paul and Susan Chik, have given me more than I can properly express thanks for. I am immensely thankful to them for supporting me even in uncertain (and sometimes apparently foolish) endeavors, and for their constant love and tireless help especially over the last several years of PhD work. My siblings, Yolanda and Joshua Chik, remain my best friends from childhood: I am grateful to them for their affection, and for many amusing and lighter moments when these were much needed. I wish to thank also David and Victoria Breidenbach, with other family and friends in

Cedarburg, Wisconsin, for their kind hospitality and encouragement over these last four years.

This thesis is dedicated to the memory of my father, Paul Chik, who died on 8 March 2014. It is also dedicated with very special appreciation to my husband, Michael, for his faithful support from the very beginning.

Table of Contents

Introduction	I
Chapter I: Disjunctivism in Action	II
1.1 Action causalism	11
1.2 ‘Caused in the right way’	16
1.3 Causal exclusion and the Cartesian background	21
2.1 Alienated or disappearing agents	30
2.2 Objections from the spontaneity of action	39
3.1 The argument from illusion	47
3.2 Action disjunctivism	57
3.3 Transitional remarks	71
Chapter II: Aristotle as Philosopher of Action	73
1.1 The possibility of an Aristotelian theory of action	73
2.1 Was Aristotle an action causalist?	77
2.2 Objections to the causalist interpretation	83
2.3 Further difficulties	87
3.1 Action as Aristotelian change	91
3.2 First difficulties: ‘change’ and natural phenomena	98
3.3 Aristotelian change as a disjunctivist approach	100
3.4 Difficulties with ‘end-directedness’	105
4.1 Action as psychophysical process	111
4.2 Examples of psychophysicality	117
4.3 Aristotle’s contemporary contribution	121
4.4 Prevention of disappearing agents	124
4.5 Difficulties in developing the account	127
4.6 Contesting the contemporary alternative	130
4.7 Non-decomposability	132

4.8 Inextricability ‘in definition’	135
Chapter III: Contemporary Adaptations of Psychophysicality in Action	140
1.1 Introduction	140
2.1 Hanna and Maeise’s ‘essential embodiment theory’: trying	142
2.2 Preliminary implications for philosophy of mind	150
2.3 Difficulties	157
3.1 The ‘essential embodiment theory: property fusion	163
3.2 Property fusion as a contemporary alternative: initial concerns	168
3.3 Complementarity in action explanations	171
3.4 Hierarchy and layered worlds	177
4.1 The causal exceptionalism of action	180
4.2 Agent causes and causal pluralism	185
4.3 Criticisms, and an integrative approach	197
Conclusion	215
Bibliography	216

Introduction

The standard story of action, also known as ‘the causal theory’, or ‘causalism’, is familiar to readers of contemporary philosophy of action. Established in the decades following the publication of Davidson’s seminal essays, the standard theory promotes a concept of action as constituted by a bodily event, joined to certain mental conditions by a bond of causation. Because the standard or causal theory’s definition of action is in terms of the components of action, the theory is properly understood as adopting a reductive approach. The causal theory is ‘reductive’, that is, in its presumption that analysis of behavioral events can and should begin at the level of the constituent elements of those events. The causal theory holds another important feature of interest, one that is perhaps related to its reductive approach. Philosophers have observed that causalists conceive of the primary causal relation as being placed ‘between events or states of affairs or facts of which [the agent] is *a constituent*’ (Broadie 2013: 574, italics mine).¹ Because the causal theory’s attitude towards the agent is that she too is a constituent of the causal relata productive of action (i.e., the events of an agent *S*’s intending to *A* and *S*’s *A*-ing, where *A* is a set of movements belonging to an agent *S*), the agent herself is conceptually divided into constituent parts of distinct types: she is ‘split’, for purposes of the causal analysis, as a constituent of events or facts of a mental type, and constituent of events or facts of a physical type. Thus the

¹ This description of the causal theory is explicated in fuller detail in the first chapter (§1.1). The notion of ‘being a constituent’ at work here expresses merely a logical relation between causal relata identified as events, states of affairs, or facts, and the agent herself. Therefore, that an agent may be described as a ‘constituent part’ of events, states of affairs, or facts, should not, of course, imply that the sense of parthood employed here is anything like material parthood with respect to a whole. Such use of the term ‘constituent’ rather emphasizes the causalist implication that the agent’s involvement in such entities is explanatorily secondary to the events, states of affairs, or facts, whose status as the primary causal relata of action is maintained by the causal theory of action.

causal theory can also be described as ‘post-Cartesian’ in its treatment of mental and physical phenomena.² The latter ‘post-Cartesian’ idea and its related reductive approach together constitute two assumptions that are central to the causal theory of action.

This thesis argues that both of these core assumptions should be resisted. Its aim is not only to provide reasons against accepting the causal theory of action (by rejecting its core assumptions), but also to propose an intelligible alternative to the standard view in the philosophy of action. This alternative claims two broad influences in particular. First, it aspires to develop the idea of ‘disjunctivism’ as applied to the concept of action. ‘Disjunctivism’ can be understood as representing a philosophical movement that has attracted followers in several diverse areas of philosophy, although it primarily holds interest for philosophers of perception. There is disagreement among the latter group as to how ‘disjunctivism’ ought to be defined, in addition to significant varieties in its application (e.g., metaphysical, epistemological, and experiential disjunctivism each hold a distinct position and address diverse subject matters). This thesis, however, simplifies the terms of the discussion by assuming that disjunctivism about perception entails, at minimum, the idea that experiences can be divided into two fundamentally different kinds, i.e., veridical experiences, and non-veridical (illusory or hallucinatory) experiences. Following recent work on the range of this idea, this thesis also adopts the insight that ‘the mark of disjunctivism, in all of its varieties, is a conception of the inner and the outer as suffused’ (Haddock and Macpherson 2011: 22). The metaphor of inner and outer ‘suffusion’ provides a

² That the causal theory is sometimes described as ‘post-Cartesian’ does not typically arise from any conception of the causal theorist as holding a commitment to the Cartesian metaphysics of substance dualism. Very few philosophers would defend the latter, so it is unlikely (and in most cases would be explicitly denied) that Cartesian *substance* dualism plays any significant role in forming the causal theorist’s conception of action. Nonetheless, a ‘post-Cartesian’ framework has relevance in discussions of mind and action insofar as a fundamental division between mental and physical types is commonly assumed to have explanatory power.

preliminary thesis for one variant of disjunctivism about action, which differentiates between action and mere bodily movements (e.g., the involuntary or non-voluntary). Attempts to develop this variant of disjunctivism are thus proposed and tested throughout this thesis. Given that disjunctivism still retains a relatively minority position among its standardly causal counterparts, it is hoped that the thesis also contributes to a growing area of scholarship on disjunctivism in relation to action theory.

The other major influence of this thesis is Aristotle. Aristotle's thought bears no immediate resemblance to present-day debates concerning 'the problem of action', which is typically characterized by the familiar Wittgensteinian question, 'What is the difference between my arm's going up and my raising my arm?'³ Yet arguably no other ancient thinker has held so much influence in shaping the field of contemporary action theory. Causal theorists⁴ and non-causalists⁵ alike identify certain features of Aristotelianism as providing critical components of their respective views. Whether Aristotle can be legitimately claimed as a precursor to *both* is surely an intriguing historical question, but the answer to it has implications for philosophy beyond the historical.⁶ This thesis will argue that the causal theory's claim to Aristotle as the

³ Wittgenstein's particular formulation of this question arises in *Philosophical Investigations*, §621: 'Let us not forget this: when 'I raise my arm', my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?' (2001: 136).

⁴ Such as Aguilar and Buckareff (2010), who name Aristotle as a precursor to the causal theory of action. Davidson also writes: '[T]he best argument for a scheme like Aristotle's is that it alone promises to give an account of the 'mysterious connection' between reasons and actions' (1980: 11).

⁵ Such as Anscombe (2000). Stoutland remarks: '*Intention* is permeated with Aristotelianism' (2011: 6).

⁶ An important point to consider with regards to this historical-philosophical question is that Aristotle's ideas about causality and motion are very different from those assumed by contemporary discussions. By that fact alone it will be difficult to claim Aristotle as a precursor for either contemporary causal or non-causal theories. One relevant feature of Aristotle's account, which becomes a focal point of the third

progenitor of its view cannot be substantiated. Instead, there exist good reasons to think that Aristotle's treatment of mind and action bear significant compatibility with the idea of disjunctivism about action, or at least a specific variant of it. With regards to this latter hypothesis, the thesis considers two interpretations of Aristotle's thought on action in particular. The first addresses his concept of action as change, which has been used to argue that representation of the philosophical difference between action and mere movement relies on invoking the notions of Aristotelian change and non-change. The second interpretation of Aristotle considers his concept of action as a unified psychophysical process. The latter concept, it will be argued, provides a substantial basis for the disjunctivist claim that action and mere movement are fundamentally different kinds of behavioral occurrences, while also promoting the further disjunctivist insight that the differentiation between these kinds rests on a metaphorical idea of 'suffusion' between the inner and outer aspects, or the psychological and physical processes that occur simultaneously when agents act.

The apparently historical considerations of the thesis are therefore justified by the hypothesis that Aristotle's idea of psychophysical process potentially assists in further developing a variant of the contemporary disjunctivist concept of action. Aristotelian psychophysical processes can be broadly interpreted as occurring in natural contexts outside of human agency: for the majority of non-human animals, too, exhibit a characteristic agency that can be described as psychophysical. Thus this variant of disjunctivism about action holds the additional advantage of appealing to a certain kind of naturalism. The final objective of this thesis, then, is to focus on developing the Aristotelian variant of disjunctivism about action in plausible contemporary terms. A recent account of psychophysicality, known as 'embodiment theory' (Hanna and

chapter, is his claim that causes (in paradigm cases) are not events, states of affairs, facts, or properties, but rather constituents of these, i.e., agents and the matter on which they act. For the majority of contemporary views, the causal locus is assumed to exist between events, states of affairs, facts, or properties, rather than in the constituents of these.

Maiese 2009), is considered, with the conclusion that the intelligibility of the latter account depends on appeal to a recent variant of top-down causation (Steward 2012).

Before proceeding to the chapter summaries, some explanation is needed as to why the thesis is titled ‘The Unity of Action’. Admittedly, this title may present a somewhat opaque description of my aims. However, I believe that its broadness may also be evocative in suggesting several areas of ‘unity’ that the thesis desires to achieve. As I have suggested, one of its central concerns is the post-Cartesian framework presumed by contemporary discussants concerning mind, agency, science, and naturalism. If agents are analytically divided as constituents of events or facts of a mental type, and constituents of events or facts of a physical type, then the problem of causal exclusion persists, i.e., of how the efficacy of mental events can be established in physical domains. A metaphysical gap appears as well in the post-Cartesian concept of action, between the latter’s causal relata; some philosophers have described it as a ‘sort of vitiating gap – whether metaphysical, temporal, or both – between conscious intentionality and its immediate bodily effects’ (Hanna and Maiese 2009: 105). The causal theory of perception presents yet another form of disconnectedness in its conception of the causal relation between the perceiving subject and the mind-independent object. The sense datum version of the causal theory is particularly illustrative of the fact that, on the post-Cartesian framework, the perceiving subject is denied access to her physical (mind-independent) environment, for the physical object of her awareness is always mediated via a mental phantasm⁷, whether her experience is

⁷ And in some extreme versions of the causal account, the subject’s awareness is always primarily of a mental object or phantasm. Clearly, on this latter picture, the disconnectedness between the subject and mind-independent objects is exacerbated further. I am not claiming, however, that all causal accounts maintaining a notion of mental phantasms as mediating necessarily result in a picture of subjects as ‘disconnected’ from physical reality, even if only to a lesser extent than the extreme version of the causal account suggests. I would not deny, that is, that there are plausible causal accounts in which the problem of disconnectedness does not arise. But I do not review these anywhere in my thesis, for my discussion of perception has

veridical or not. Yet another example of disunity may be drawn from Davidson's argument that the capacity for articulating reasons provides the only basis for identifying one's movements as rational, and therefore, by extension, for marking it as intentional; thus non-human animals, on Davidson's account, are incapable of action. This latter view has been the source of much trouble for the philosophy of action, as it promotes a picture of agency as requiring propositional speech or propositional mental content, separated from the 'merely' animal movements that are obviously shared by both rational and non-rational agents.

The Aristotelian variant of disjunctivism considered in this thesis attempts to overcome or avoid entirely each of these problematic views. The 'unity of action' describes the psychophysical unity of agents and their intentional doings, natural phenomena that are not exclusive to human beings. The concept of psychophysicality corresponds to the simple idea that agents are embodied beings, integrated in mind and body, with capacities likewise that exhibit this integration of both their physical and psychological aspects, simultaneously and irreducibly. A conclusion of the thesis is therefore that such psychophysical unity is an unsurprising and widespread feature of the natural world. The thesis also concludes that consideration of the concept of an animal agent makes it entirely philosophically unparadoxical that the mental and physical are always found together in voluntary movement, and that the embodiment theory's central notion of 'property fusion' potentially complements a naturalistic variant of top-down causation in explanations of agency.

The summaries of the chapters follow.

In the first chapter, I define and present the case for 'action causalism'. Particular attention is given to the Davidsonian variant, partly out of regard for its prominence,

the limited purpose of illustrating an analogy between action and perception, i.e., an analogy for which the more extreme versions of the causal theory enable me to draw a more useful comparison (Chapter 1).

but also simply for reasons of space. Several objections to action causalism are then considered, along with the causalist replies to each. The first objection arises from the widely cited problem of deviant causal chains. The second objection, adapted from the philosophy of mind, involves the problem of causal exclusion, which, I argue, persists in light of a widely accepted Cartesian argument linking language and mental causation in agency. Third, causalism is frequently said to promote alienation between action and agent, and I consider a novel version of this objection based on close analysis of the concept of basic action. Fourth, reflection on the frequently spontaneous character of action leads to the conclusion that causalism appears to be phenomenologically false. Fifth, causalism faces problems analogous to those arising from the argument from illusion, which promotes a causal view of perception. Both action causalism and the causal theory of perception problematically entail a ‘common factor element’ that disables the perceiver or agent from having, respectively, access to her environment or (in the case of action) control over what happens. Finally, a disjunctivist approach for perception is sketched as having potential utility for the case of agency. The conclusion of the first chapter is that further work is needed to fully appreciate the insight that disjunctivism potentially holds for action theory. In particular, I propose examining the ways in which Aristotle’s thought has influenced both causalist and non-causalist proponents, and whether an interpretation of his concept of action can be usefully employed to develop the disjunctivist approach to action.

The second chapter therefore engages with possible diverse interpretations of Aristotle as a philosopher of action, after acknowledging some initial doubts over whether such an undertaking is even possible. My inquiry is divided into two general parts. First, I examine the case for thinking that Aristotle was a kind of action causalist. Causal theorists, beginning with Donald Davidson, frequently claim Aristotle as the progenitor of their own views, and I examine the sources for substantiating this claim. Having found the case for making it unjustifiable, I examine

a moderately contrasting position that recognizes the demerits of the causal theory while defending the idea that Aristotle accepted some very generalized causalist theses. This approach focuses on a concept of action as Aristotelian change, while also suggesting an interpretation of Aristotle as the inspiration for modern disjunctivist approaches to action. I consider several significant deficiencies of the latter interpretation of Aristotle, with the conclusion that the problems with the account outweigh its advantages. The second general part of my inquiry therefore considers the more nuanced view of action as a psychophysical or hylomorphic process. The overall aim of this chapter is to consider the extent to which Aristotle's concept of action is able to vindicate the disjunctivist approach to action, which previously was deemed lacking in substance. Its conclusion is that Aristotle's concept of psychophysical process can be intelligibly defended as being compatible with and in support of a certain variant of the disjunctivist approach to action.

Finally, in the third chapter, I consider a contemporary approach to articulating the concept of action as psychophysical process. The overall aim of this chapter is to refute skeptics of Aristotelian psychophysicality who believe that there is no credible theory of action that can improve on causalism using the notion of psychophysicality. I consider a recent approach offered by Hanna and Maiese's *Embodied Minds in Action*, which offers a contemporary Aristotelian version of a 'non-classical causal theory of action' (2009: 157). Hanna and Maiese's argument for 'trying' in action has the compelling advantage of avoiding deviant causal chains, but a number of their assumptions are problematic. Their proposal of property fusion holds great interest, but it remains unclear whether their overall model succeeds in distinguishing itself from other solutions, especially supervenience theories in general. In addition, the approach to action defended by proponents of 'trying' is not incompatible with a variant of agent causation recently proposed by Steward, and I suggest that a specific agent causal view that spells out animal agency as top-down control over the parts of the subvenient base, can be viewed as an extension of the psychophysical composite

account inspired by Hanna and Maiese. Particular attention is paid to the fact that the variant of top-down causation considered does not render the causation of events, properties, and facts irrelevant, nor conceive of these entities as inefficacious. My thesis agrees with Steward's approach in claiming that the causation by substances (e.g., agents) is explanatorily primary to all other causal types that may potentially be involved in action explanation.

This thesis attempts to advance the contemporary debate in at least the following ways. First, it proposes a more serious contemporary rapprochement with Aristotle's thought on action. Aspects of Aristotelianism, as I have mentioned, are in one way or another often appealed to by philosophers of action, but the appeal mostly relies on a very generalized and superficial level of interpretation. My thesis attempts to engage more carefully with Aristotle's position, and in doing so it therefore counters some overly facile interpretations or appropriations of Aristotle, while granting credibility and support to others. Second, the thesis endorses a disjunctivist way of thinking about the concept of action, and thus attempts to build upon the recent body of literature on disjunctivism. The latter represents a growing area of scholarship, but disjunctivism remains a minority position, especially within the philosophy of action. My thesis endeavors to show that a disjunctivist approach to action is an intelligible alternative, which holds appeal especially for philosophers of action who identify as Aristotelian and/or as non-causalist. Finally, the thesis adds to recent discussion on the idea of agent causation and naturalistic explanation, by defending a view of agent cause as explanatorily prior to all other causal varieties.

This thesis is by no means comprehensive. There are particularly difficult areas that will need to be investigated more fully before the thesis may be adopted with full confidence. For instance, the rapprochement between Aristotle and contemporary philosophy may be charged with holding anachronistic assumptions, since Aristotle's conceptual framework, his areas of interest, and his scientific perspective, are all

significantly at odds with the contemporary philosophical scene. Non-Aristotelians may doubt that integrating his thought with a disjunctivist approach holds much relevance, while Aristotelians themselves may complain that his concepts are wrongly characterized or applied ahistorically (and therefore amount to an egregious account for other reasons). Regarding the latter complaint, I have wherever possible sought for accuracy in my treatment of Aristotle, although my exposition is surely not free from error. The more difficult achievement, I believe, lies in providing a compelling account of how Aristotle's concepts might be usefully introduced in contemporary discussions. The interpretation of action as a psychophysical process is my starting point, but some may see the thesis's consideration of a recent contemporary development of psychophysicality ('property fusion') as somewhat of a false step, given the many problems associated with the concept of property fusion. In any case, my hope is that my critical treatment of this latter account reveals some important mistakes and false assumptions that a contemporary Aristotelian would do well to avoid, with the conclusion that there are other routes to follow in developing the concept of psychophysicality. Even if the reader cannot have, as yet, full confidence in the view advanced by this thesis, it aspires to be at least a very promising working hypothesis.

Finally, this thesis presents several areas of inquiry to which one might naturally turn next. I will mention only a few of them. For instance, further varieties of action disjunctivism might be explored, in addition to the disjunctivism based on psychophysical process emphasized by the thesis. An area left open by the thesis concerns the possible connections between a disjunctivism involving the agent's reasons for acting and the disjunctivism between action and mere movement explored by the thesis. There is also a question of how the thesis of 'animalism', the view that human beings are essentially animals (or, more strongly, that we could not exist except as animals), might be considered an extension of the idea that agents are causally unexceptional in the natural world, and that agents also include non-human animals.

Chapter 1: Disjunctivism in Action

The case for ‘action causalism’ is laid out, with particular attention to the Davidsonian variant. Several objections to action causalism are then presented. The first is the problem of deviant causal chains. Second, causal exclusion remains a problem for causalists, but persists in light of a widely accepted Cartesian argument linking language and mental causation in agency. Third, causalism is frequently said to promote alienation between action and agent. Fourth, reflection on the frequent spontaneity of action leads to the conclusion that causalism appears to be phenomenologically false. Fifth, causalism faces problems analogous to difficulties that arise from the argument from illusion, which promotes a causal theory of perception: both action causalism and the causal theory of perception problematically entail a ‘common factor element’ that disables the perceiver or agent from having, respectively, direct access to her environment or direct control over what happens. Finally, a disjunctivist approach for perception is sketched as having potential utility for the case of agency.

1.1 Action causalism

The prevailing account of action in the contemporary action theory literature is known variously as ‘the causal theory of action’ (Aguilar and Buckareff 2010), ‘the standard story of action’ (Smith 2012), and ‘action causalism’ or, simply, ‘causalism’ (Wilson 1989). Aguilar and Buckareff define the causal theory of action (CTA) in the following terms:

(CTA) Any behavioral event *A* of an agent *S* is an action if and only if *S*’s *A*-ing is caused in the right way and causally explained by some appropriate

nonactional mental item(s) that mediate or constitute *S*'s reasons for *A*-ing (2010: 1).

This definition of (CTA) is perhaps deliberately broad. It allows for a variety of different conceptions of the mental causes of *S*'s *A*-ing, while following the Davidsonian principle that such mental items always constitute *S*'s reasons for *A*-ing. The broadness of Aguilar and Buckareff's (CTA) also leaves open several critical questions. Foremost among them is the issue of what being 'caused in the right way' requires, and in particular, whether it requires more than merely the 'appropriate nonactional mental items' that constitute the agent's reasons for acting. (CTA) also raises the fundamental question of how a mental item can cause a physical event. This latter problem is understood by philosophers of mind as the problem of 'causal exclusion'.

Another problem that arises from the definition of (CTA) is that Aguilar and Buckareff's descriptions of a 'behavioral event *A* of an agent *S*' and '*S*'s *A*-ing' are used as if interchangeably. Such usage, however, overlooks an important difference between them: a 'behavioral event *A* of an agent *S*' appears to entail that the event *A* belongs to an agent *S*, whereas '*S*'s *A*-ing' places emphasis on an event of which the agent *S* is a logical 'constituent' or element.⁸ These formulations are, in fact, incompatible with one another. The former description suggests that events belong to agents, with the potential implication that agents are the primary causes of action (in which events may be a part or aspect), whereas Davidsonian causalism assumes that

⁸ As clarified in the Introduction, such use of the term 'constituent' expresses a logical relation between agent and event, comparable to the way that 'the vase is on the table' may be said to have the logical constituents of 'the vase', the 'being on' relation, and 'the table'. Perhaps this is an overly loose way of describing the relation between agent and events (or states of affairs, or facts), but it should be clear that the constituent relation employed here bears no significant similarities (except analogously) to metaphysical parthood, e.g., as when the legs of the table are understood as its constituent parts, among others.

the relevant causal relata are events, of which agents are merely logical constituents and therefore secondary in explanatory relevance. Where action is described in terms of the agent, to whom behavioral events belong or are attributed, there may be hope that a sense of the agent as primary cause of her own movements may be retained.⁹ In contrast, action descriptions of the form ‘*S*’s *A*-ing’, in which any mention of *S*’s involvement might be easily left out so as to produce abstract descriptions such as ‘walking’, ‘cooking’, and ‘writing’, seem to emphasize the following implication of the causal theory: that the agent’s involvement in such entities is explanatorily secondary to the status of events, states of affairs, or facts as the primary causal relata of action. Even if the causalist believes that one form of action description, ‘a behavioral event of an agent *S*’, is interchangeable with or reduces to the other, ‘*S*’s *A*-ing’, it cannot be assumed that such interchangeability or reductive identity holds likewise of the respective conceptions of the causal relata of action, which may be derived from these distinct forms of action explanation: i.e., (i) the agent as primary cause of her own movements, and (ii) the agent as a logical part or mere explanatory addendum to events (or states of affairs, or facts), where these, rather than the agent, hold primary causal status.

The potential conceptual differences between these descriptions of action must therefore be addressed and minimized, if the causal theorist is to be justified in using these respective descriptions interchangeably. One notable problem that seems to arise naturally from the conceptual discrepancy just considered is the ‘problem of the disappearing agent’ (§2.1 of this chapter), which many philosophers believe to pose a serious objection to the causal theory of action. One may worry that the formulation ‘*S*’s *A*-ing’ makes too easily available certain abstract descriptions of action (e.g., ‘walking’), in which the causal relevance of an *agent*, i.e., as principle source or

⁹ Arguments for this claim will be considered in Chapter 3.

performer of the action, is rendered otiose.¹⁰ In addition, ignoring the differences between the two conceptions of agents and events leads to further difficulties in the way ‘reasons’ are conceived: whether, as on Davidson’s view, reasons are states of mind (a pro-attitude and belief ‘bundle’), the occurrence of which cause the event of *S*’s *A*-ing, or whether it is more natural to think of reasons as the agent’s considerations in favor of doing something or not. These questions will need to be answered if (CTA) is to have any plausible standing.¹¹

According to Smith’s ‘standard story of action’, the difference between *S*’s being an agent (doing something, *A*-ing, etc.) versus a patient (merely being involved in something’s happening), ‘lies in the causal etiology of what happens when a body moves’ (2012: 387).¹² Smith suggests two criteria for distinguishing an action from a

¹⁰ This problem plays a significant role in the thesis, with considerations of recent articulations of the disappearing agent worry considered in §2.1 of Chapter 1. One of the conclusions of that discussion, and in the thesis overall, is that if this particular objection to the causal theory is to have any merit, it will need to be made on other grounds than on those presented by Velleman, Hornsby, Lavin et al. In particular, a disjunctivist proposal that incorporates the Aristotelian concept of psychophysicality and a recent variant of top-down causation will be defended as securing the primary causal role of the agent.

¹¹ The Davidsonian interpretation of CTA also raises an important question that this thesis leaves aside: this is the issue of whether the putative mental causes (generally understood) of *S*’s *A*-ing are intelligibly equated with *S*’s reasons (conceived as situational features in favor or against) for *A*-ing. Dancy, for example, argues against this version of CTA (2002). The interest of this thesis lies far less in the nature of reasons and their various types (although the latter is a worthy topic on its own), than in the character of certain mental features of action that other philosophers may assume have to do with an agent’s ‘reasons’.

¹² Hanna and Maiese characterize the causal theory of action in the same way, with the further objection that causal theories of action are committed to an element in common between what an agent does versus what merely happens. Thus they write: ‘Classical causal theories of action say that what distinguishes intentional actions from mere bodily events is essentially a difference in the causal origin of those bodily events. The bodily events are the same in both cases, and intentional actions are supposed to be the events brought about by antecedent causes in some categorically

mere movement. First, the action must be a ‘basic action’ in Danto’s sense: it is something that the agent is able to do straightaway, without his knowledge how to do it being mediated by the knowledge how to do something else (Danto 1963, 1965).¹³ Smith’s justification for this criterion of basic action follows from his methodology, for he begins by stating that one identifies any putative action ‘by tracing its effects back to some bodily movement’: if the bodily movement is an action, then it must be something that the agent can do straightaway, without the mediation of knowledge how to do something else (2012: 387). Smith’s second criterion for establishing whether the agent acts is ‘by seeing whether this bodily movement [the basic action] is caused and rationalized in the right kind of way by some desire he has that things be a certain way, and some belief he has that a basic action of his – specifically, his moving his body in the way under discussion – has a suitable chance of making things the way he desires them to be: that is, he has to have a relevant means-end belief (2012: 387). Smith’s approach to the causal theory of action differs slightly from the Davidsonian view, on which events are the causal relata: the latter account holds that the agent *S*’s having a desire and relevant belief is the event that causes *S*’s *A*-ing, whereas Smith’s account claims that *S*’s desire and relevant belief are mental *states*, which cause *S*’s *A*-ing.¹⁴

different manner’ (2009: 103). This common factor feature of causal theories is the topic for the later sections of this chapter, especially §3.1.

¹³ Chisholm proposes a similar view to Danto’s. He describes a basic action as the ‘undertak[ing of] certain things *p* without undertaking certain other things *q* for the purpose of bringing about *p*. The things *p* that we thus undertake could be said to be things we *undertake directly*. When we are successful with respect to the things we undertake directly, we may be said to perform certain *basic actions*. These are the things we succeed in doing without undertaking still other things to get them done’ (1976: 209).

¹⁴ In all likelihood, the discrepancy is not a very significant one, especially for causalists who believe that states can and should always be recast in the form of events (from ‘the state *p* causes’ to ‘*S*’s having the state *p* causes’).

Smith's definition of 'the standard story of action' faces nearly the exact same set of questions as arise in Aguilar and Buckareff's (CTA). The familiar qualification, 'caused and rationalized in the right kind of way', obviously remains an issue needing clarification. Smith's definition introduces a further question concerning 'basic action' and its conceptual connection to the standard story. Lavin has argued against the acceptability of this latter connection: although seeming to be 'a gift of common sense', the legitimacy of the notion of basic action is one that should be treated with skepticism, despite its being 'vital to the intelligibility of the *causal theory of action*, according to which physical action consists of a mere event and a condition of mind joined (in the right way) by the bond of causality' (2012: 2).¹⁵ This problem, too, will need to be addressed before the proponents of causalism can rest assured that their account of action is indeed the best available story.

This chapter is intended to consider the principal objections to action causalism that arise from these questions. Some of these objections, especially the problem of deviant causal chains, have received very prolonged attention over the last several decades, and for these I will do little more than summarize their main points, rather than offer an exhaustive account. A less widely considered challenge to action causalism draws from areas of comparison to the causal theory of perception, and the latter part of the chapter will be occupied with reflecting on the questions that arise from this comparison, and sketching an alternative 'disjunctivism' that may be plausibly applied to the case of action.

1.2 'Caused in the right way'

The causalist's qualification that a movement is an action *iff* that movement is 'caused in the right way' produces an ambiguity, as already mentioned. The most considerable difficulty that arises from this ambiguity is the problem of deviant causal chains. The causal theory of action claims that, for *S*'s *A*-ing to count as an action, the event *A*

¹⁵ Lavin's argument will be considered in greater detail in §2.1 of this chapter.

need only be caused by appropriate mental states of the agent *S*. The problem of causal deviance is that, for *S*'s *A*-ing to count as an action, causation of a bodily movement by the appropriate mental states does not always seem sufficient.

Davidson's seminal example of the rock climber illustrates the problem famously:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally (1963: 79).

In this case, the sufficiency of certain appropriate mental items – the belief and the desire appropriate to *intentionally* loosening one's hold on the rope – is lacking as far as the anticipated action is concerned. The appropriate mental states occur, and they play a causal part in producing a bodily movement, but that movement is not an action. Despite meeting the causalist's basic requirement for action, i.e., that a mere movement be joined causally to a condition of mind, what happens to the climber remains a mere movement. So it appears that more is needed, as a sufficient condition for the occurrence of the kind of event understood as action, than merely the mental causes claimed by the causalist to rationalize and produce the relevant bodily movement.

Something has gone wrong in the account of the causation of action: the requisite mental items occurred, but they did not cause a bodily movement 'in the right kind of way', for deviant or wayward causal chains intervened between the occurrence of the appropriate mental states and the consequent bodily movement. The causal chains normally leading to action have instead run wayward, *because* the appropriate mental items did not cause movement 'in the right way': but what precisely is the right kind of way? Smith argues: 'The obvious response to this problem [of wayward causal

chains] is to note that, in cases of internal wayward causal chains like that described, the match between what the agent does and the content of her desires and beliefs is entirely fluky' (2012: 398). Smith suggests that the notion of 'differential sensitivity', developed by Peacocke (1979), successfully eliminates the potential intervention of such causal flukes. Differential sensitivity is a further 'condition of non-flukiness' since it requires that *S*'s *A*-ing is an action if and only if *S*'s *A*-ing is differentially sensitive to the contents of *S*'s desires and beliefs, so that had the agent *S* have had a slightly different set of psychological circumstances, *S* would have acted in response to the contents of *those* beliefs and desires (2012: 398). Smith explains:

[W]hat the differential sensitivity requirement guarantees is that whenever an agent acts intentionally, he doesn't just try to realize the desires he actually has, given the means-end beliefs he actually has, but that he would have tried to realize his desires, given his means-end beliefs, in a range of nearby possible worlds in which he had desires and means-end beliefs with ever-so-slightly different contents (2012: 398).

Peacocke and Smith's notion of counterfactual sensitivity therefore entails responsiveness to psychological *content*, and not merely to the occurrence of a suitable belief and desire pair. Similar 'sensitivity' arguments have also been suggested, sometimes along the lines of 'proximate' or 'immediate causes' of actions, in order to close the gap between intention and action into which deviant causal chains might be inserted.¹⁶

¹⁶ See, for instance, Schlosser (2007). Schlosser casts doubt on the idea that an intention or some 'reason state' is always the 'immediate cause' of an action, since the causal chains between intention and action apparently always run through non-intentional, physical states, or are realized at (or determined by) the neurophysiological level (2007: 192-193).

Despite acknowledged difficulties¹⁷ surrounding the condition of counterfactual sensitivity, Smith concludes that its overall approach is successful. Other philosophers disagree. Steward, for instance, writes critically of the approach: ‘The trouble with most versions of this idea...is that counterfactual sensitivity to reasons, intentions, and the like may be present in a situation in which, intuitively, the agent still fails to act’ (2012: 57). She imagines a hypothetical in which an artificial neurophysiological intervention produces exactly the kind of motor impulses that lead to bodily movements corresponding to an agent’s known (but impotently realizable) intentions. The agent possesses the capacity to form beliefs and desires, and therefore may have intentions, but the connection between her mental states and her motor impulses has been disabled by the neurophysiologist. The physical result therefore cannot be claimed to be action by the agent, since it is rather a movement produced artificially by a neurophysiologist, and yet counterfactual sensitivity to reasons on the part of the agent appears to be present all along. Steward concludes, ‘There must, then, it is inferred, be something more to agent control than mere counterfactual sensitivity’ (2012: 58).¹⁸ So the presence of counterfactual sensitivity to reasons (in a range of nearby possible worlds) does not, combined with the causalist’s primary requirement that suitable mental states occur, guarantee that the agent’s movement is an action.

Perhaps a causalist like Smith would reject the scenario provided by Steward on the basis that a third requirement of the causalist’s approach is missing: the agent’s

¹⁷ See, for instance, Sehon (2005).

¹⁸ Steward addresses the further condition that is often made by causalists, ‘that the causal chain from mental states to bodily movements “may not run through the intentions of another person”, as Peacocke puts it’ (2012: 58). This further condition, combined with the differential sensitivity stipulation, she identifies as a general tendency in the literature, found in accounts of the ‘classic bipartite type’. There are a number of difficulties with the addition of this further condition, primary of which is the objection that disallowing the causal chain from running through the intentions of a second agent may sometimes exclude ‘good’ cases of agency. The causalist’s reference to ‘no other agent’ in his definition of agency also produces an explanatory circularity.

capacity to act has been disabled in Steward's case, rendering it impossible that the agent's behavioral event can ever actually be *caused in the right way* by her mental states, i.e., providing that the differential sensitivity requirement to the content of her mental states is met. But this kind of reply will not do, for a few reasons. First, the causalist's account claims to be a true reductive account: it aims to provide the sufficient conditions for an agent *S*'s *A*-ing without invoking either of the notions of agency or action. For this very reason, the causalist's stipulation that an agent possess the capacity for her action to be 'caused in the right way' cannot even partially constitute a reductive explanation of what it is for *S*'s *A*-ing to count as an action: the capacity for such causation, it appears, is the very item that needs to be explained. Second, the sensitivity condition is designed to sufficiently answer the question of what action 'caused in the right way' precisely entails: the causalist's solution is that had the agent had a slightly different set of psychological circumstances (slightly differing beliefs and desires in nearby possible worlds), she would have acted in response to the contents of those circumstances. The argument from 'proximate' or 'immediate' causes employs a similar strategy: had the psychological circumstances been different, the agent's responsiveness to them would have entailed a suitable motor connection between the proximate neurophysiological causes of action and the agent's intentional movement. But Steward's scenario rules these conditions out as sufficient by themselves: the proximate cause may be present, along with the suitable belief and desire pair, and yet the agent may fail to produce an instance of true agency.

The explanatory sufficiency of the sensitivity condition thus found lacking, we may wonder if the arguments for it (and for similar proximate causal approaches) really constitute substantial progress for the original causalist thesis, i.e., that action merely requires suitable causation by mental states. Surely a significant reason that the problem with causal deviance arises is that the purported linkage between the mental items and a bodily movement fails: instead, between them there is a gap wherein

deviant causal chains might be inserted. Positing that the gap may be filled, providing that counterfactual sensitivity conditions are in place, or that proximate (to the action) neurophysiological events occur, seems like an obvious solution to the problem of causal deviance. The notion of sensitivity to psychological content is certainly an attractive one, and is rightly regarded as an improvement upon the original causalist thesis that requires merely *that* a suitable desiring and believing pair occurred. All the same, the general approach invoking a counterfactual intermediary between intention and action only pushes the problem further and further back, to the point at which what the agent does and the presence of counterfactual conditions are joined. If the agent still fails to act, then the gap remains, and the problem of causal deviance has not been satisfactorily resolved. Indeed, more than a decade after introducing deviant causal chains into the discussion, Davidson himself concluded: ‘Several clever philosophers have tried to show how to eliminate the deviant causal chains, but I remain convinced that the concepts of event, cause and intention are inadequate to account for intentional action’ (2004: 106). The inadequacy of these concepts for easing the trouble of causal deviance is, however, but one of the problems surrounding the causalist definition.

1.3 Causal exclusion and the Cartesian background

In addition to diagnosing the causal theory as failing to provide a sufficient account of the causation involved in effecting action, one should consider the suggestion that the causal relation is entirely misplaced between the relata cited. Broadie offers this further critique of the causal theory, with the charge that it ‘locates the causation in the wrong place’ (2013: 574). There could be a variety of ways to understand a criticism of this kind: opponents of the causal theory of action, for instance, have pointed out as problematic the placing of a causal locus between events (Steward 2012, Lowe 2010). Broadie’s criticism specifically targets the causalist premise that the causal relation is placed ‘between events or states of affairs or facts of which [the agent] is *a constituent*’

(2013: 574, italics mine).¹⁹ This point highlights, *inter alia*, a noteworthy but under-examined detail of the causalist's story: that the agent, understood as a logical constituent of the causal relata productive of action (i.e., the events of *S*'s intending to *A* and *S*'s *A*-ing, where *A* is a set of movements belonging to an agent *S*), implies an ontological partitioning not only between events of the mental and physical type, but also between the corresponding constituent parts of the agent herself. The causalist account simply assumes that the agent can be thus divided, as a logical constituent of events or facts of a mental type, and constituent of events or facts of a physical type. It should be noted that there is not necessarily any objection *in general* with this latter conception, e.g., as in cases where one wishes to speak of 'the event of *S*'s re-entering the atmosphere' and 'the event of *S*'s blood oxygen levels increasing', as simultaneous occurrences. The division of an agent as constituent of events is problematic specifically where one type of event is said to cause the other, as I shall consider below. This point is distinct, one should also note, from both the criticism about event causation and the disappearing agent worry, which will be addressed later.

There are at least two points of interest concerning the ontological partitioning described here. First, the division of the agent into constituents of diverse type events is clearly related to the problem of how to explain the efficacy of mental events in physical domains, known in the philosophy of mind as the problem of causal exclusion. The problem arises partly from reflection on the principle of causal closure, that every physical effect must have a complete physical cause. Thus a bodily movement must have a physical cause, but the mental items required by the causal theory of action lack, by themselves, the kind of efficacy presumed to generate physical causes. If one views intentions or reasons as metaphysically 'determined' or 'realized' by corresponding physical states, such as neurophysiological occurrences in the brain, then causal closure can be satisfied by positing that the bodily movement is caused by the physical realizers of mental events. But this kind of solution has the cost

¹⁹ See footnotes 1 and 8 for clarification on this usage of the term 'constituent'.

of literally ‘excluding’ mental events, such as intentions or reasons, from the causal story. They are rendered epiphenomenal, or causally inefficacious, and may be dismissed as mere artifacts of ‘folk psychology’. But this is a cost that relatively few causalists are willing to accept.²⁰

The second point of interest concerns the nature of the break that occurs, whenever there is an action, between the causal relata cited by the causalist. Hanna and Maiese describe it as a ‘sort of vitiating gap – whether metaphysical, temporal, or both between conscious intentionality and its immediate bodily effects’ (2009: 105). The causalist account involves recognition of certain *types* of entities that qualify as such relata in the causation of action. Certainly this qualification based on types includes the Cartesian thought that one of them must be mental (or intentional), the other physical (or bodily), even if the way these binary concepts are understood in contemporary accounts are not exactly along the original Cartesian lines. For causalists such as Davidson (1985), the recognition of causal relata in action as being of these dual types includes the further stipulation that whatever mental or intentional state occurs, it must be of the *rational* kind in order for any ensuing behavioral event to count as an action. He offers an argument beginning with the premise that any creature with the capacity of truly possessing beliefs must also possess the capacity of exercising propositional speech. Since non-human animals are incapable of expressing propositional speech, they must (by the first premise) be incapable of truly possessing beliefs. And since belief possession is a requirement for rational action, according to the causalist’s basic definition, Davidson concludes that non-human animals are therefore incapable of performing actions, which by his lights represent an event of the *rational* type.²¹

²⁰ An alternative that this thesis addresses later is identity theory, which attempts to resolve the problem of epiphenomena by positing that mental states just are, no more and no less, brain states.

²¹ McDowell offers a similar argument, on Kantian grounds (1996).

What is the connection, if any, between Davidson's argument from speech as one condition for action, and the recognition of binary event types as another? It should be noted that there may not be a necessary link between these conditions, at least not one explicitly stipulated by causal theories, whether of the Davidsonian variety or some other kind. All the same, it is of more than mere historical interest to recall that the original Cartesian argument not only gave the same argument from the capacity for speech as a condition for mind and action, but it also linked this argument to the recognition of binary event types: that each should be drawn from the realms demarcated by *res cogitans*, the thinking or mental substance, and *res extensa*, the extended or physical substance. For Descartes, the connection between such substance dualism and a creature's propensity for propositional speech was one of simple logical entailment: it followed from his view of physical substance as being, for the most part, separately bounded from mind. Declarative or propositional speech, he wrote, 'is the only certain sign of thought hidden in a body' (1649/1991: 244-245). Although non-human animals exceed human beings in certain abilities, their lack of inventiveness in respect of declarative speech, Descartes argued,

proves that they do not have a mind, and that it is nature²² that acts in them, according to the disposition of their organs – just as one sees that a clock made only of wheels and springs can count the hours and measure time more accurately than we can with all our powers of reflective deliberation (1637/1998: 59).²³

²² One of the issues that I will address more in the third chapter is the obvious difficulty about Descartes's concept of 'nature' that is assumed by him in this passage, as i.e., purely physical or physiological nature.

²³ This proof against the mindedness of non-human animals was definitive for Descartes. He rejected any consideration by degrees, i.e. the thought that non-human animals may be *less* minded than human beings and yet still possess some mental capacity, in his argument that the inability to communicate by tongue proves 'not

Davidsonian causalists of action may wish, for obvious reasons, to resist the absolute conclusion that non-human animals lack minds altogether, or the assertion that they are constituted much like the mechanistic assembly of Descartes's clock. Such causalists may point out that whereas Descartes draws the stronger conclusion that non-human animals' inability to speak entails the absolute absence of mind, Davidson claims that such inability merely implies the absence of belief formation and possession. It may be that Davidson's account does not outright deny some notion of mind to non-human animals, but the exclusion of non-human animals from the class of 'true' agents is no less difficult for any variety of naturalists to accept.²⁴

The most obvious problem with excluding non-human animals from the class of agents is how implausible such a claim seems in the face of current knowledge about animal behavior and learning.²⁵ Observation of the group hunting patterns exhibited by *Orcinus orca*, the largest species of oceanic dolphin known commonly as the orca whale, suggests that learnt, collaborative strategy is instrumental in species survival. The evidence also suggests that hunting and dietary tactics are not universal to all varieties of *Orcinus orca*: different groups, or 'pods', specialize in different hunting strategies and hold distinct preferences regarding prey. Such group-specific specialization suggests that the behavioral patterns are learnt over generations of the group, rather than expressions of an innate feature of the species. For example, one

merely...the fact that animals have less reason than men but that they have none at all' (1637/1998: 33).

²⁴ Some philosophers who equate 'agency' with 'morally responsible agency' in all cases will also for obvious reasons deny that non-human animals have a membership in the class of true agents. They too will face the challenge that such a view flaunts naturalistic thinking about the place of human agents.

²⁵ One might include in considerations on this topic the scientific evidence suggesting human evolution from 'lower' animal forms. An important implication of such evidence is that, if the present evidence does confirm the theory of macroevolution, naturally we should see a certain observable continuity from humans to other animal species in aspects of behavior and learning. Such continuity can be granted without at all dispensing with the idea that human beings are themselves exceptional in the animal kingdom, in obvious (also observable) ways.

well-documented hunting tactic has the goal of tipping ice floes, upon which the targeted prey has been spotted, by the group's swimming quickly in coordinated formation towards the floe, and then, in perfect unison, sharply diverging at just the right moment to create a flush of water that tips the prey into the water. There are many other examples of complex social, goal-directed behavior across a variety of animal species, fascinating to observe and far too many to list in detail. It seems truly implausible to claim that any of these instances fail to count as examples of agency.

Another problem related to the rejection of non-human animals as agents lies in its implication that human agency exhibits a form of causation that is unique or exceptional in nature. If the causation of human action is unique among the causal phenomena found in nature, then nothing about non-human animal behavior can be informative with regard to human behavior. I will say more about this problem of causal exceptionalism in the latter chapters of this thesis, but it suffices for now to note that the claim of action as causally exceptional among natural phenomena will again be deeply unattractive to any variety of naturalists about human beings.

A third difficulty with Davidson's claim concerns its implications for what counts as human agency. Causalists will not deny that non-human animals engage in some kind of *movement*, even if such movement falls short of qualifying as action. According to the Davidsonian causal theory, the movement of non-human animals fails to be intentional, because of the absence of beliefs whose contents are propositional. This means that, for a movement to be intentional (i.e., for the movement to count as an action), it must be caused, *inter alia*, by a psychological event or state with certain declarative or propositional qualities or contents. The picture of human agency that follows has *mere movement* on one side, where such movement is uncaused by such a cognitive state or event, and true action (intentional movement) where the latter cause is present.

To see why this picture of human agency is problematic, it is helpful to revisit a passage from the *Philosophical Investigations* that has been treated over the last decades as an important starting point for philosophical thinking about action. Wittgenstein asks:

Let us not forget this: when ‘I raise my arm’, my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm? (2001: 136).

The question is often presumed to ask for an explanation of the difference between what agents do and what merely happens to them. There may be cases in which one’s arm goes up without the agent having raised it: perhaps the agent suffers from ‘alien’ or ‘anarchic’ hand syndrome, a neurological disorder (usually caused by brain injury or surgical separation of left and right brain hemispheres) that causes autonomous, ‘rogue’ limb behavior. These cases are obviously distinct from cases in which the agent, having full and healthy control over her limbs, raises her arm, and yet it seems that they also have something in common: in both cases there is a bodily movement.²⁶ They may even look and feel the same: imagine a scenario in which a neurophysiologist has disabled my ability to control my limb movements, and connected them artificially so that, when my arm is made to go up, the sensation is identical to one in which I have absolute control over my body. The causalist’s explanation of what distinguishes them lies solely in the causal etiology leading to the behavioral event. As earlier considerations showed, the Davidsonian variety of the causal theory insisted that such etiology include cognitive states or events with specifically propositional or declarative content. The first difficulty with this picture lies, again, in its implausibility. Davidson’s causal picture suggests that human activity

²⁶ There is an important issue that arises here, one relating to the fact that in both cases there is a bodily movement, which will be addressed later (the problem of a ‘common factor’ view).

may be comprised of many movements, but only those that have a suitable causal etiology count as intentional and agential. This account, responds Steward,

has always faced a particular sort of embarrassment. The embarrassment is that a great deal of our purposive activity does not really seem to be preceded at all by the sorts of mental events and states that figure in the story. Often, I just seem to *act*, without any prior deliberation, without first deciding or choosing to do anything, and without at any stage consciously forming an intention to act in the way that I do, or being conscious of the existence of such a prior intention (2012: 66).

To the causalist, Steward issues the sound advice of taking notice, for just ten minutes, the variety of bodily movements that may occur. These likely include crossing and uncrossing of the legs, touching one's face, stretching, etc., not to mention the 'scratchings, shufflings, twiddlings, and jiggings that together constitute quite a large sub-class of our bodily movings [that] present the most obvious counterexamples to the claim that to act is to have a bodily movement caused by something like a prior intention, decision, or choice' (2012: 66). Causalists will argue that these counterexamples are not conclusive. Steward counters that such attempts are inevitably 'tortuous'; and, in any case, 'what exactly is the reason for insisting on preserving it?' (2012: 67). Such insistence on preserving the basic causalist idea of action's specific mental antecedents 'must be a reason for suspicion that the...model...is basically Cartesian. The agent is identified with certain of her mental states and events, and her settling how her body will move is then thought of as *their* settling how her body will move; that is to say, in what looks from this point of view as though it ought to be the best case, as their deterministically causing the wanted bodily movements. But we humans are not purely mental beings, we are embodied ones' (2012: 67). This last statement is one that my thesis has the aim of rigorously developing, especially in the third chapter. The point at present is that Davidson's

causalist account *is* basically Cartesian, particularly in its solution to Wittgenstein's infamous question. For Descartes before him gave essentially the same explanation of action in the Sixth Meditation:

I might regard a man's body as a kind of mechanism that is outfitted with and composed of bones, nerves, muscles, veins, blood and skin in such a way that, even if no mind existed in it, the man's body would still exhibit all the same motions that are in it now except for those motions that proceed either from a command of the will or, consequently, from the mind (1641/1998: 101).

The above passage is notable not only because the metaphysical independence of mind apart from the body is clearly asserted. It also vividly illustrates a view of matter belonging to the agent (not merely the body but also the bodily motions and behavioral events exhibited by that body) as a kind of underlying substratum for intentional movement. As for the contemporary causalist, what differentiates action from mere movement is simply the addition of a mental antecedent that figures into the causal etiology of the behavioral event: 'a command of the will or...the mind'.

It is hopefully clear from the foregoing that Davidson's distinctively Cartesian approach has little to do with Descartes's metaphysical doctrine of substance dualism. The dualism assumed by Davidsonian causalism is rather one concerned with sets or systems of concepts. One set of concepts is employed for describing matter, and another, very different set is used for describing mind; neither conceptual system makes reference to any of the concepts contained in the other. I have suggested in this section that the problem arising from this view, the problem of causal exclusion, produces difficulties for the idea of a mental event or state as causing the event of a physical movement. That the latter idea encounters difficulties, in the case of contemporary causal theories, is not usually the fault of any *metaphysical* picture in which ghostly or incorporeal substances play a putative role in moving or pushing

around physical entities (unless, of course, one really does subscribe to Descartes's metaphysical dualism). Instead, the difficulty of accepting physical causation by mental events is an *explanatory* difficulty, for the physical domain is believed to be wholly explicable in purely physical terms.²⁷

2.1 Alienated or disappearing agents

Besides the problem of deviant causal chains, the most frequent criticism of the causal theory of action is that it promotes a concept of the agent as 'alienated' from his action. The most influential critiques along these lines have been offered by Velleman and Hornsby. Velleman introduced the issue by characterizing the role of the agent in the causal theory of action as 'the disappearing agent' (1992). The problem, according to his critique, is that 'of finding an agent at work amid the working of the mind' (1992: 131). The causal theory of action 'fails to include an agent or, more precisely, fails to cast the agent in his proper role. In this story, reasons cause an intention and an intention causes bodily movement, but nobody – that is, no person – *does* anything' (1992: 189). Velleman also clarifies that his objection to the causal story 'is not that it mentions mental occurrences in the agent instead of the agent himself [but] that the occurrences it mentions in the agent are no more than occurrences in him, because their involvement in an action does not add up to the agent's being involved' (1992: 463).

But it appears that the causalist has an obvious reply to this type of objection. Being a reductive account, the causal theory's very aim is to cite occurrences below the level of the agent in its explanation of action. On these terms, it would be recursive to include the agent in its explanation. Yet it does not follow from the fact that sub-agential occurrences are instead cited, that the agent fails to be involved. Why indeed should

²⁷ A complementary question might arise, although I do not address it anywhere in the thesis: why is it not a problem (for materialist proponents who accept explanatory exclusion of the mental) that a physical change or event might cause a mental event, e.g., a physiological change that causes an agent to become aware of that change?

it? On the causalist's account, the sub-agential occurrences are no less occurrences of the agent, a sub-agential exemplification of her ability to perform actions. I consider this causalist reply again in the second chapter.

Hornsby's complaint against the causalist begins similarly, but contains differences from Velleman's charge. She writes: 'The idea of an agent at work goes entirely missing from Davidson's story of human agency. Davidson spoke of causality as "central to the concept of agency", but he said that "it is ordinary causality between events that is relevant, and it concerns the effects...of actions"' (2011: 108). Hornsby objects to the causalist idea that the causation of action involves a relation between events (or states) rather than agents, and offers as counterexamples a list of transitive verbs, such as 'carry', 'push', and 'squash', which are of the kind in which their passive form ('being carried', 'being pushed', 'being squashed') 'cannot be pried apart' from the active form that includes the agent as subject, such as *S*'s 'carrying' or 'pushing' an object *o*. By stating that these forms 'cannot be pried apart', Hornsby suggests that there exists no possible gap between agent and action in which causal intermediaries, such as events or states, can be inserted, such that '*S* pushes *o*' cannot be rendered as "the event of *o*'s being pushed" is caused by "the event of *S*'s having an intention (or belief and desire) to push *o*". Descriptions such as these, Hornsby argues, involve the use of events as 'mediating' between their constituents, *S* and *o*. She remarks: 'We might say that in such cases the person does the thing – carry something, or, as it might be, push, or squash something – *non-mediate*ly' (2011: 108). Hornsby's problem with the causal theory, it appears, is that the latter fails to countenance a schema in which agency is mediated by anything other than events or states.

It may be that Hornsby's objection is subject to the same causalist reply as was addressed to Velleman's concern, but there is another interesting feature of her objection to consider. This is the fact that Hornsby chooses to center her critique on

the notion of action without intermediaries. If it is true that the causalist requires the mediation of events or states for the causation of action, then there may indeed be a problem about whether *the agent* mediates them, and if so, how? If the agent does not mediate them, then how do they arise, and how then could it be said that the agent is truly ‘at work’? Admittedly, there is also some potential confusion over the concept of mediation as Hornsby employs it. She appears to use the term ‘mediation’ simply wherever events are taken to be the causal relata, so that an ‘unmediated’ action expresses a concept of agent (rather than event) as cause.²⁸ But events and states are not typically conceived by causalists as playing a sort of mediating role, in any conventional sense of mediating: rather, these are cited as the reductive components of a complex that needs to be analysed, such as ‘*S* pushes *o*’.

Thus, the former questions raised by Hornsby’s account may be of practical irrelevance if the causalist does *not* require a concept of action as mediated at all.²⁹ Lavin, for example, argues for the seemingly opposite conclusion that action must be properly understood as ‘mediated’, if agents are not to be forced into a position of alienation from their actions. Lavin’s notion of ‘mediation’ is undoubtedly very different from Hornsby’s, although both offer arguments against the causal theory as promoting a concept of the agent as alienated from her actions. Lavin argues that the concept of ‘basic action’, specifically an *unmediated* event that is done ‘just like that’,

²⁸ Perhaps, on the other hand, Hornsby’s concept of ‘mediation’ is meant to play a role in our understanding of actions: i.e., that we can understand ‘*S* pushes or squashes *o*’ without *our understanding* being mediated by the assumption that the causal relata involved in the description of *S* and *o* are mental events and physical movements. I admit that it remains unclear, at least on my reading of her, whether an ‘unmediated action’ entails this interpretation or something else. Either way, however, Hornsby’s remarks amount to little more than a refusal to accept the causal theory of action. It may be that she simply wishes to remind us of the naturalness of thinking in agent causation terms, in which ‘the person does the thing’, rather than a causal relation between events.

²⁹ That is not to say, however, that Hornsby’s objection against event or state causation assumed by the causalist is of practical irrelevance: merely that her usage of the concept of ‘mediation’ is arguably too loose to correspond precisely to that objection.

is ‘vital to the intelligibility of the causal theory of action’ (2012: 2-3). Danto’s original concept sought to reflect the structure of a basic idea: he described basic actions as ‘not compounded out of anything more elementary than themselves, but [are] instead the ultimately simple elements out of which [actions are] compounded’ (1965: 147). According to Lavin, an implication of Danto’s ‘practical atomism’ is that basic action is therefore ‘barren of means-end structure’ (2012: 2). What this means, he explains, is that the performance of basic actions is never mediated ‘by any such thought thought about how to bridge the gap between here, where one has not yet done what one aims to do, and there, where one has brought things to completion. With basic action, thought about how to realize the end in question gives out – one acts immediately, directly or “just like that”’ (2012: 4).

Lavin extends two principal claims concerning this definition of basic action. First, he explains that the notion of basic action is central to contemporary analyses of action, for both standard theories (i.e., the causal theory of action) and non-standard theories alike (e.g., agent causation approaches).³⁰ Second, Lavin argues that the contemporary concept of basic action cannot in fact sustain all the conceptual work that contemporary theorists lay upon it: for ‘basic action’, if his argument succeeds, turns out not to be action at all.

He begins with the first point by explaining the nature of the assumption of basic action held in common between both causal theorists and non-standard approaches alike:

³⁰ Lavin cites Smith, a causal theorist, who defines actions according to the standard story: ‘Actions are those bodily movements that are caused and rationalized by a pair of mental states: a desire for some end, where ends can be thought of as ways the world could be, and a belief that something the agent *can just do*, namely, move her body in the way to be explained, has some suitable chance of making the world the relevant way’ (2004: 165, italics Lavin’s). Lavin also cites Bishop, who describes the proponent of agent causation as also requiring an unmediated event: ‘All intentional actions have a component which is a *basic action*, and it is here that the agent-causalist professes to find an irreducible causal relation between agent and event’ (1983: 62).

‘[W]hatever the specific disputes, the parties to them share a generic conception of a material process or event (the conditions of identity and individuation are free of intentionality) and the explanatory ambition of fitting action into a world of material processes so understood. And thus they share an allegiance to the very general framework of the causal theory of action: that *X* did *A* intentionally is the arithmetic sum of *what merely happens* and something else’ (2012: 7).

Basic action is unmediated because it ‘is not the end of any other action; nothing else is done in order to do it; it is not an answer to “Why?” when asked about any other action’ (2012: 3-4). Lavin argues that the causal theorist in particular needs to isolate an event that lacks this sort of intrinsic intentionality³¹ in order to make good on the thesis that action consists of a mere event joined by a condition of mind. A closer analysis of Davidson’s example of turning on the light reveals a regressive series: Jones turned on the light via his intention to turn on the light, but the causalist’s basic causal relation depends on a further connection: Jones’s flipping the switch resulted in the light turning on (2012: 8). But Jones’s flipping the switch is yet another intentional item, so the causalist has not yet successfully isolated the necessary constituent of his equation, the ‘not-intrinsically-intentional event whose causation in the right way can constitute intentional action’ (2012: 8-9). Lavin concludes:

Basic action is what stops this regress. If ‘*X* did *A*’ reports a basic action, then, *ex hypothesi*, *X*’s doing *A* does *not* involve *X*’s having done anything else intentionally. Here, finally, we have an action which might itself be identified with non-intentional movement caused by thought. Here, we

³¹ It could be added that the kind of ‘intrinsic intentionality’ lacking in basic action (as Lavin wishes to conceive the latter) not only does not offer an answer to the question ‘Why?’ but also precludes the agent’s own question ‘How?’. For if an action is basic, according to Danto’s definition, then one does not need to ask ‘How?’ in doing it: there is no prior action that the agent must take in order to perform a basic action, that would necessitate the question ‘How?’.

reach a practical atom, something that might be shown to stand as the correlate of a mere happening. Of course, with the basic case in place the causal theory will extend the analysis by describing how non-basic action ultimately decomposes into an aggregate of basic actions, caused by intentions with such-and-such contents, which can themselves be decomposed into independently intelligible, metaphysically independent, internal and external elements, the mind of the acting subject and what merely happens (2012: 9).

The concept of basic action thus expresses a fundamental commitment to a notion of material processes as metaphysically independent of the mental. Complex (non-basic) action, understood as intentional movement, requires ‘atomic’ decomposition in order to reflect this fundamental commitment. Lavin’s argument is that the causal theory depends, for the sake of the causalist arithmetic, on the existence of basic action ‘as the correlate of a mere happening’. For the causal theorist shares the assumption that ‘thought or reason must stop short of movement itself: it cannot reach all the way into the constitution of what happens. If the jurisdiction of practical thought must be so limited, so must the jurisdiction of specifically instrumental thought, the rational ordering of means to ends – there must be basic action’ (2012: 9). If Lavin is right, then his analysis discharges the general thrust of (one interpretation of) Hornsby’s criticism: that the causal theory implicates the alienation of an agent from her action by interposing causal intermediaries. Lavin does not wish to dispute that the causal theory implicates a form of alienation between the agent and her action, however. His argument for this, which is yet to follow, is that such alienation results rather because the causal theorist insists on the simplest constituent of action as being entirely without intermediaries. For the causal theorist’s most critical assumption, he argues, is that basic action exists: as an atomic constituent of an agent’s intentional movements, a mere happening isolated from intention and thought, to which certain

conditions of mind may be causally joined (in the right way) to produce action. Thus his primary aim is to cast doubt on the truth of this idea.

Consideration of the implications of this assumption leads to Lavin's second and more ambitious aim, which is to show that the notion of basic action *cannot* exist and function as it does according to the general contemporary analysis of the structure of action.³² Lavin begins by arguing that certain features of contemporary discourse tend to obscure reflection on the nature of action and its temporal structure. In particular, 'the tremendous reliance on abstract noun phrases ("action", "event") and name-like phrases ("X's doing A to Y", "Y's A-ing") in the formulation of general philosophical theses conceal the bad fit with most any pretheoretical characterization of the phenomenon of which a theory is wanted' (2012: 19). Lavin proposes that philosophers focus on the aspectual distinctions between the schematic sentences, 'X is doing A', 'X was doing A', and 'X did A', and as represented concretely in action descriptions such as 'I am walking across the street', 'I was walking across the street', and 'I walked across the street'. The aspectual opposition between the perfective and the imperfective suggests a sense of completion and incompleteness, of something that may at one time be still underway, and at another is underway no more. Thus 'the poles of the aspectual opposition are not simply at odds. Here *doing* looks forward to *having done*: when walking across the street is in progress, it is not then finished, but nevertheless it reaches ahead to completion' (2012: 19-20). The imperfective 'reaches

³² Lavin offers a substantial preliminary discussion in support of a weaker claim, that basic action *need* not exist in consequence of certain regress arguments that might variously be given by contemporary theorists (see 2012: 9-17). These regress arguments may take different approaches, but for instance, one may believe that there must be simple skills if there are complex ones. Such belief may correspond to Hornsby's argument, as cited by Lavin, that '[a]mong the things a person knows how to do, some of them he must know how to do "just like that", on pain of needing to ascribe to him indefinitely many distinct pieces of knowledge to account for his ability' (1980: 88). Lavin's discussion of the regress arguments is worth considering, but for present purposes I focus only on his stronger argument that basic action *cannot* exist and function as contemporary action theorists believe that it does.

ahead to completion’ even in cases where the walker is interrupted, for it ‘has intrinsic direction; it is a specification of what is to be, even if not what will be in fact’ (2012: 20). The aspectual opposition so conceived clearly draws on Aristotle’s concept of *kinēsis*, a movement that is performed to completion over time.³³

Lavin’s purpose with regard to these considerations is to develop an argument against basic action from them, i.e., that ‘when described in light of this temporal structure basic action will not be recognizable as action’ (2012: 21). His argument begins with the premise that the temporal structure of processes so described, as following Aristotle’s concept of *kinēsis*, is true. He proceeds:

1. Suppose *X* did *A*, which is a basic action.
2. From the temporal structure of processes, it follows that earlier *X* was doing *A* and had not yet done it.
3. It also follows that at the time *X* is doing *A*, other things have already happened, i.e., *X* did *A*^{**}, and at the moment that *X* is doing *A*, still others are underway, i.e., *X* is doing *A*^{*}. The latter subordinate phase can be subject to the same temporal analysis as ‘*X* is doing *A*’.³⁴
4. Since doing *A* is basic, it follows that any subordinate phases to *A* do not themselves involve anything intentional, for they ‘are not themselves undertaken in pursuit of the goal’ (2012: 22).
5. Therefore, the subordinate phases that constitute the progress of *A*’s being done ‘towards its completion is thus wholly opaque to its subject, except in the way it might be known to an observer or to someone with general knowledge of

³³ Lavin cites Aristotle at *Nic. Eth.* 1174a 22-24: ‘In their parts and during the time they occupy, all movements are incomplete, and are different in kind from the whole movement and from each other’.

³⁴ ‘And so, earlier when *X* was doing *A*^{***} and had not yet done it, other things had already happened in progress (*X* is doing *A*) is at once an ever increasing stack of *have done*’s and ever shrinking list of *still to do*’s’ (2012: 22).

how such things happen. In short, the subject of basic action is alienated from the progress of her deed' (2012: 22).

Lavin argues that the subject of basic action, with regard to the progress of her deed, stands in the same position as one who is unaware of the unfolding of a causally dependent action, i.e., one that need only be set in motion to progress independently of the agent's awareness or will. Davidson's example of the naval officer who sinks the Bismarck illustrates the concept of a causally dependent action aptly. Having launched the torpedo (perhaps by merely pushing a button), the naval officer relinquishes control of his intended deed, which progresses even after his death. 'Mere movements of the body', Davidson stated, 'are all the actions there are.... We never do more than move our bodies: the rest is up to nature' (1971: 59).³⁵ Lavin's argument in the above steps (1)–(5) is intended to show that basic action is itself a similar kind of *kinēsis*, the nature of whose progress is comparable to the progress of causally dependent actions such as sinking the Bismarck. The completion of the latter may be satisfied entirely outside of the naval officer's control and awareness, so long as he pushes the proper buttons. Likewise, the completion of a basic action, defined as a mere movement in which there is no intrinsic intentionality, apparently proceeds entirely outside of the intentionality and awareness of the one who somehow triggers it (since, by definition, basic action lacks rational thought and thus also intentionality). Precisely because the

³⁵ In another paper, Davidson considers a different situation in which his right arm is paralyzed, but it is placed in a pulley system wherein he can raise it by pulling on a rope with his left arm. 'Raising an arm is usually done without doing anything else, but not always', he comments (2004: 103). In this case, Davidson claims that he has raised his paralyzed right arm by pulling on the rope with the other. Although raising an arm is usually something that one can do 'just like that', he has accomplished it only by doing something else with his body, an extra causal intermediary. Having pulled on a rope in order to raise his other arm, however, Davidson denies that achieving this latter end is part of the action at all: '[T]he answer to Wittgenstein's question for a case like this is that nothing is added to the rising of my arm that makes it a case of my raising my arm because the rising of my arm is not part of my action at all' (2004: 104).

movement is basic, it cannot be constituted by subordinate phases any of which themselves are intentional or possessed of rational thought. Thus, the concept of basic action renders the agent superfluous with regards to the subsequent unfolding of sub-events that constitute the basic event.

Lavin remarks: ‘It looks like performing a basic action is just being the subject of a mindless, automatic process which the subject has somehow initiated, triggered, or launched’ (2012: 23). Having triggered the automatic sequence, the agent is relegated to a position of passive observer of what happens, much like Marx’s *Nichtarbeiter*, or non-worker, who is ‘alienated from the material processes that realize his own ends and fortunes’ (2012: 18). Although the non-worker may have awareness of these material processes, ‘such awareness would be of something alien, outside and external, something given through observation – a matter of being put into contact with what is already there’ (2012: 18). Likewise, such is the condition of the agent, if Lavin’s argument is persuasive: under the assumptions of the causal theory of action, alienation of the agent from her action is a consequence of putatively basic action failing to be action at all, for there is no subordinate part of its progress that is guided or purposive.

2.2 Objections from the spontaneity of action

To escape these charges, the causal theorist can either probe for weaknesses in Lavin’s argument in steps (1)–(5), or entirely jettison the notion of basic action so understood, i.e., as critically underpinning the causal theory of action via the ‘practical atomism’ described by Lavin. It is unlikely that the latter alternative will be attractive to the causalist, since it undermines the causal theory’s fundamental structure, i.e., that an action is constituted by a mere movement joined by certain mental states. If the notion of basic action were to be made unavailable for this thesis, then the causal theory would be forced to revise its claim that action is understood in terms of the causal conjunction of mental states and *basic* movement, i.e., movement that lacks

rational thought and intentionality. But this would then require the causalist to define action circularly, in terms of something already intrinsically intentional. The causalist must therefore find a defense against Lavin's argument, whose weak point is arguably the transition from steps (4) to (5). One might argue that the lack of intrinsic intentionality in basic action should not in all cases entail that the agent is faced by utter opacity with regard to her movements. Even among certain involuntary or non-intentional movements, which by definition lack intentionality, subjects can sometimes claim awareness of what happens to their bodies: of one's involuntary kick when the knee is tapped, for example. But the causalist is still burdened to show that basic action is movement of this kind, i.e., lacking in intrinsic intentionality and yet something of which the agent may be aware. In any case, it remains unclear how the progress of such movements can ever exist as transparent processes to their owners, i.e., as subjects of the non-observational type of awareness that seems essential to action.

Lavin's argument represents a novel critique of the causal theory, but there may be doubts about his analysis even from the non-causalist perspective. For instance, in the previous section (§1.3), I mentioned a number of counterexamples to the causalist's claim that to act is to have a bodily movement caused by something like a prior intention, decision, or choice. Based on these considerations, surely a rejection of causalism would entail that there *are* some things that we do 'just like that', spontaneously: i.e., without having to identify by deliberation a means to doing it? It seems that Lavin would deny this important feature, which arguably seems to apply to the majority of our voluntary bodily movements. For he argues:

The capacity to perform a basic action 'boils down to a capacity rational agents have of getting [something] done without needing to cognitively control how it is done.' The passage is from Berent Enç, who is admirably clear-sighted about the need for this sort of commitment. To

Enç it seems ‘intuitively clear’ that we have this capacity. To me it seems intuitively clear that we do not have an intelligible conception of rational agency, if we just have something that announces the thing to do, leaving the execution of the task to another power. Reason is not practical (i.e. efficacious) if it is simply agenda setting in this way: it must also be sufficient to the realization of its ends, and thus must be able to constitute the progress of the deed, the getting done of what gets done (2012: 23).

Lavin may be right about these latter commitments to the concept of rational agency, but it is not clear that Enç’s remark really entails what Lavin assumes of it. Indeed there appears to be an ambiguity about what is meant by ‘cognitive control’. The term may imply the sort of conscious guidance of one’s movements that is minimally expected of agents in the paradigm cases of intentional activity (e.g., baking a cake, walking to school, swimming laps in the pool). Construed differently, ‘cognitive control’ could also imply that an agent must deliberate about the means – the ‘how’ and perhaps also the ‘why’ – of what she would like to achieve. The latter construal could include the requirement that mental antecedents must be in place for the agent to do anything intentionally. If Enç assumes the latter definition of ‘cognitive control’, then he is surely right that rational agents (and, one should add, non-rational agents as well) have the capacity of doing something ‘without needing to cognitively control how it is done’. Why indeed should this plausible proposition entail a conception of agency on which one simply ‘announces the thing to do, leaving the execution of the task to another power’?

Lavin’s argument apparently relies on the first construal of ‘cognitive control’, the denial of which would be absurd even for the most unreflective or mundane of activities: for even occasionally stirring the soup in the pot requires an amount of guidance from the agent, throughout the moments in which one is stirring. Lavin’s

protest is against a view that would deny instrumental means-end structure to seemingly mundane movements like stirring the contents of the soup pot once around. For to reiterate his argument discussed in the previous section – if stirring the contents of the pot once around is a basic action in the sense that there is *nothing* cognitive involved at all, then the subordinate phases of stirring the contents once around (i.e., stirring the contents half-way around, stirring them a quarter of the way around, and so on) likewise cannot involve anything cognitive on the part of the agent. If so, then it appears that the progress of such basic action is unknown to (and alienated from) the agent: for to know about one’s own movements, in the non-observational sense, surely requires some sort of cognitive involvement.

The potential difficulty in Lavin’s analysis lies in whether the causal theorist’s denial of ‘cognitive control’ implies that basic action involves *nothing* cognitive at all. The causal theorist may well protest that Lavin simply assumes a notion of basic action whose primary characteristic is its absence of rationality and intentionality, i.e., an event that can hardly be identified as ‘action’ at all. As already noted, one might presume (as Eng appears to presume) that ‘cognitive control’ denotes deliberation on the means of what the agent sets out to do: in which case one’s denial of the latter would hardly entail the thesis that the agent’s movements entirely lack cognitive involvement at all. As for the construal of ‘cognitive control’ as agential guidance and awareness of her own movements, it is hard to believe that the causalist would really insist on a conception of basic action as something that agents *do* and yet are utterly unguided movements of which they are wholly unaware.³⁶ It seems clear enough that

³⁶ By the same token, Lavin’s assumption that action is either ‘bound together by instrumental rationalization’ or wholly lacking in rationality altogether, could be criticized on the grounds that he cannot conceive of rational action as ever falling in between these polar opposites (2012: 13). However, it may be that his conception of ‘instrumental rationalization’ is so broad as to admit of actions that one does as a matter of routine or habit, if not ‘just like that’. It will be difficult to see the coherent application of ‘instrumental rationalization’ to spontaneous or habitual actions *if* one assumes that ‘instrumental rationalization’ involves cognitive control of the

the definition of (CTA), the causal theory of action (§1.1 of this chapter), implicates a concept of ‘mere’ movement empty of rationality where ‘rationality’ is presumably some variant of instrumental rationality, but not agential awareness *tout court* only for the purposes of defending the causalist’s particular ‘atomistic’ approach to defining action. ‘Basic action’, on this view, is a term of art that enters only later into the analysis of complex action, i.e., as involving several steps or stages, versus simple or basic ones.

This is not to say, of course, that the causal theory of action may not be criticized for its reliance on an atomistic approach to explanation of action. One may object, for instance, that atomic analyses of action incompletely or inaccurately characterize action from a phenomenal standpoint. The foregoing discussion on cognitive control raises this further objection for the causal theory from the perspective of actions done spontaneously, i.e., without the agent’s having to deliberate about how to do it. Having distinguished between different definitions of ‘cognitive control’, it becomes possible to accept Enç’s concept of basic action without also committing the causalist error that Lavin identifies as promoting alienation between the agent and his action. Basic action, depending on one’s construal of ‘cognitive control’, may indeed be understood as one way to oppose the ‘atomistic’ tendency of the causal theory: i.e., the thesis that action requires decomposition for its analysis, with the line of decomposition falling between purely material processes, or mere movement, and thought or reason, understood as metaphysically independent of the material processes in which the action consists. Broadly understood, basic or simple actions are rational but also unreflective, failing the mental states requirement of (CTA), for they specify movements that are intentional and yet uncaused by any identifiable event of mind. Such actions are not subject to ‘atomistic’ analysis, in that they lack

deliberative kind, i.e., requiring the agent to deliberate on the means of what is to be done. A broad conception is conceivable only if the latter assumption is thrown out, e.g., as is arguably the case with Anscombe’s particular understanding of practical knowledge, with which Lavin appears to sympathize.

decomposable elements. They themselves are non-decomposable. This, at least, is a possible tentative rejoinder to standard theories of action, which presume that distinct elements of action can be separately identified and analysed.³⁷

This objection from spontaneous action will surely be resisted by the causal theorist, on the grounds that certain mere empirical facts cannot disprove that spontaneous actions are not also caused by distinct mental states. For my present argument appears to rely on the idea that since spontaneous actions are not preceded by (or co-occurrent with) an agent's deliberation about how to perform them, one cannot easily identify any distinct mental states that might empirically play a causal role in the agent's movements. In contrast, for actions that are preceded by deliberation (or where deliberation is experienced concurrently with the action), it is easy to identify distinct mental states that can be said to cause one's movements. Lacking this feature of co-occurrent or antecedent deliberation, spontaneous actions therefore may be presented as empirical counterexamples to the causal theory of action. The causal theorist will certainly reply³⁸ that distinct mental states are present in spontaneous action nonetheless, as the cause(s) of one's movement. But there is no reason to accept the causalist's rejoinder here other than for the purposes of upholding the causal theory, for the causalist's reply arguably amounts to a case of denying the facts for the sake of one's theory. The causalist begs the question in favor of his theory if only on the basis of his theory does he maintain that the empirical facts are as the theory dictates.

The Aristotelian framework is one obvious candidate for resisting the atomistic tendency of the standard analysis. Within it there can be action that is done

³⁷ This sort of reply is not incompatible with Lavin's account, since he countenances actions that are done unreflectively, as a matter of habit or a part of routine activity. His account, however, specifies that such actions must be 'bound together by instrumental rationalization' (2012: 13).

³⁸ Michael Smith made this exact reply to me in Leuven, May 2011.

spontaneously, ‘just like that’, without much deliberation or other causally antecedent cognitive processes or states. Yet such action is not less ‘rational’, especially in the sense that it may be done for some end. In Book II of the *Physics*, Aristotle mused that it seems ‘absurd to suppose that purpose is not present because we do not observe the agent deliberating’ (199b 27). The swallow need not deliberate to build its nest, nor the spider deliberate to spin its web, he observed, and yet *their* actions are no less done for an end (199a 10-34).³⁹ Action without deliberation (either antecedent or concurrent) may also be ‘rational’ given Aristotle’s account of *skill*: his idea that, with proper training – which involves necessary deliberation and careful monitoring of one’s movements – an agent ‘internalizes’ the steps or rules of a particular skill. Thus in his public performances Heifetz required no deliberation to play his violin flawlessly, for in the many years of deliberative training, a virtuoso gains the ability to perform non-simple actions with little effort. But the capacity for spontaneous action, given its Aristotelian background, is not reserved exclusively for non-human animals and geniuses: mediocre mortals likewise perform complex tasks without deliberation, such as dicing an onion, cycling into town, and typing up long articles on a computer

³⁹ Aristotle offers a more extensive treatment of this topic in *De Motu Animalium*, especially §7. There may be a syllogism that one considers, that then leads to action: ‘I need a covering, a coat is a covering: I need a coat. What I need I ought to make, I need a coat: I make a coat. And the conclusion “I must make a coat” is an action’. Though this type of syllogistic practical reasoning reveals a general structure, it need not be the case that an agent *dwells* on such propositions, or that they even occur to him. ‘[H]ere the intellect does not stop and consider at all the one proposition, the obvious one; for example if walking is good for man, one does not dwell upon the proposition ‘I am a man’. And so what we do without reflection, we do quickly. For when a man is actually using perception or imagination or thought in relation to that for the sake of which, what he desires he does at once. For the actualizing of desire is a substitute for inquiry or thinking. I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink. In this way living creatures are impelled to move and to act, and desire is the last cause of movement, and desire arises through perception or thought imagination and thought. And things that desire to act, make and act sometimes from appetite or impulse and sometimes from wish’ (701a 6-701b 1).

keyboard. If these considerations are mounted as an objection to the causal theory, however, they admittedly cannot be viewed as a definitive one. The causalist argues that the apparent absence of deliberation from the third-person perspective does not definitively show that there is none: when asked, the human agent can surely give reasons as to why she did this or that, usually formulable into a specific desire and a suitable belief, although it is not necessary to cite both. Swallows and spiders, of course, cannot give reasons: thus they are excluded, along with the other non-human animals, from either deliberation of the kind specified by causalists and action itself. As I have already mentioned, however, this sort of reply seems to assert the truth of one's theory regardless of the empirical evidence.

It is also worth asking the causalist: why should it be assumed that, as a general philosophical condition, consideration of the concept of action need only engage in the analysis of action beginning at the level of decomposition, with the atomized elements of a behavioral event as its necessary and sufficient conditions? For it is possible to hold that rigorous philosophical examination of a topic need not only engage in analysis in this literal (i.e., decompositional) sense. I have no doubt as to which of my movements are my own *actions*: a reflex kick produced by the doctor's examination could not be more different from one's kick of the ball for a goal. First-person phenomenological certainty usually applies whether the action is a very complex and multi-staged one, such as building a house or baking a cake, or a spontaneous action, perhaps something done out of habit, like rubbing a statue's foot for luck as one daily enters and exits a building. An agent is certain that these are her actions even when, on occasion, there is no obvious reason for which she does them. And it does not seem unreasonable to claim that she can be certain about many of the movements of other agents, as well, many of which are recognizably actions. Why indeed should the swallow building a nest, or a spider spinning a web (the former instance in particular exhibiting complex social practices that are learned and taught

within a specific group) be excluded from the class of agential movements, the particulars of which we can be reasonably sure to count as actions?

The causal theory of action flouts our phenomenological inklings about our own actions, and the actions of others. It shifts priority away from a first-person perspectival framework so that it may, in fact, be irrelevant whether something *feels* to an agent like her own action, so long as the necessary suitable mental events are in place.⁴⁰ Phenomenological considerations by themselves may not amount to definitive objections to the causal theory, but they are very much worth raising, if only to question the explanatory priority that the causalist places on entities at the atomized level.

3.1 The argument from illusion

In the remainder of this chapter, I would like to consider a specific problem that is debated by philosophers interested in the metaphysics of perception. I believe that an analogous version of the problem arises for the causal theory of action, and likewise, that a certain response given to the perceptual problem, known as disjunctivism, can also be applied to the realm of action theory. My purpose is not to give a comprehensive treatment of the problem of perception and the reply of disjunctivism, which has multiple formulations in the literature; that would be an impossible task for the allotted space of this thesis. Instead, I want to suggest merely that certain lines of discussion from the literature on perception can be usefully drawn to form the basis

⁴⁰ One may be able to come up with counterexamples in which the necessary and relevant mental events are in place, and yet the movements that follow do not feel to me like they are my own action. (Conceivably there may be a case in which my hypothetical paralysis permits me to have necessary and relevant mental events or states; upon observing the occurrence of these states a knowledgeable neurophysiologist then makes it happen that the relevant part of my body moves. Clearly, in this situation, the movement of my body would be alien to me. What happens certainly cannot count as ‘action’.)

of a different sort of reply to the causal theory of action, which was criticized along more conventional lines in §§1-2 of this chapter.

The problem of perception can be explained as a feature of the argument from illusion.⁴ Philosophers distinguish between ‘good’ and ‘bad’ cases of seeing: those in which the perceiver has a veridical or true experience of an object, and those in which she has a hallucinatory or illusory experience. The argument from illusion takes for granted the phenomenal indiscriminability between the veridical and illusory experiences: it presumes that, whether the experience is of a true object or an illusory

⁴ Philosophers of perception have presented several diverse ways to advance an argument from illusion; the summary I offer here is a very basic version involving sense data (there could be an alternatively basic version that does not involve sense data, however). The general argument may be presented in two stages. The first part brings to bear the causal ‘immediacy’ of perceptual awareness, whether veridical or not: in light of such cases, one might conclude that the ‘immediate object’ of one’s experiences are not objects in the external material world, but rather an entity or event of a very different kind, i.e., possessing a non-material nature. John Locke and George Berkeley spoke of ‘ideas’ or ‘ideas of sense’, while Hume conceived of ‘impressions’: all of these suggestions serve as forerunners to the contemporary notion of ‘sense-data’. Their main implication is simply: that which we immediately experience or perceive is never material, but constituted by states or events that are independently mental. The general claim, then, is that our visual perception of external objects is always mediated by perception of a state or event distinct from the external objects of perception themselves. In other words, the immediate objects of perception are not external. Bonjour presents a second stage to the argument, based on the claim that ‘immediate experiences of “true” qualities are not distinguishable by appeal to their intrinsic character from immediate experiences of “false” qualities’ (2002). If we accept this possibility of introspectively (or phenomenologically) indiscriminable cases, ‘what is *immediately* experienced in both sorts of cases are objects or entities or whatever exactly they are of the same basic kind, ones that exists only in or in relation to the experience. At least in the cases involving the “true” qualities, we can also be properly said to experience the material object that really has those qualities – but not *immediately*’ (2002: 114). However, it is conceived, the argument retains the basic form of adding a further object or event to the common constituent of an appearance. For instance, Dancy describes it thus: ‘A theory of [this] sort is tempting: an appearance is genuine perception when it is caused by an external object which it sufficiently resembles. Causal theories like this are theories about what one has to add to appearance to get perception’ (1995: 422).

one, a perceiver's sensory experience has all the same qualities, except for the fact that an 'external' object is really present in one case and absent in the other.⁴² It can be inferred from this indiscriminability that the perceiver's visual experience is of qualitatively one and the same thing, regardless of its veridicality. What the perceiver is 'directly' aware of is therefore the same, regardless of veridicality: one's direct awareness is not of the external object itself, but rather a mental image or sense datum of it.⁴³

A noteworthy feature of this version of the argument from illusion is that it assumes that the mental image or sense datum, of which the perceiver is directly aware, must be some purely mental entity that is possessed in common by both the true and the deluded perceivers. The true perceiver sees what her deluded counterpart sees, with the difference that what she sees is actually present. The implication of isolating this mental item or object of perception is that veridicality is then understood as an event

⁴² The nature of illusion (or hallucination), as it appears in the argument, is not typically presumed necessarily to correspond to *real* illusions or hallucinations. Conceived as experiences that are phenomenologically indistinguishable for the perceiver from a veridical perception of an object, 'illusions' and 'hallucinations' in the argument from illusion (or hallucination) are understood to be merely possible events. Whether this assumption of *possible* experiential indiscriminability for the subject is a potential weak point for the proponent of the argument, is an issue taken up specifically by Martin (2006). Most philosophers, whether disjunctivists or not, however, accept as a common presumption that the argument from illusion is merely describing a merely possible event, i.e., experiences that *may* be indistinguishable for the subject from the real thing.

⁴³ As I suggested in one of the previous footnotes, I am perhaps over-simplifying the case by presuming the involvement of sense data. The reference to sense data need not be a critical aspect of the argument from illusion, but the argument in all of its varieties always includes some version of what Robinson (1994) calls the 'Phenomenal Principle': 'If there sensibly appears to a subject to be something which possesses a particular sensible quality then there is something of which the subject is aware which does possess that sensible quality' (1994: 32). Thus, one need not commit to the notion of sense data, while holding the view that there exists *something* of which the perceiver is aware which has the sensible quality, i.e., if only a mental state that has the appearance of the quality. For all of these qualifications, however, I believe that my main point holds.

constituted by the necessary conditions of (i) the presence of a mental object or sense datum, and (ii) the corresponding physical object itself. Because the presence of a mental object or sense datum is a common factor for both veridical and illusory experiences, the argument from illusion may be most basically understood as promoting a ‘common factor’ view.⁴⁴

The most obvious problem with the notion of a common factor in perceptual experience is that its understanding of the causal relation between perceiver and the external object denies an open link. The perceiver is denied direct access to her physical environment, for the object of her perception is always a mental phantasm, whether her experience is veridical or not. Even if the causal mediation between the mental object of her perceptual awareness and its physical correlate can be clarified, the role of the perceiver remains as from a position ‘out of touch’, or as one alienated from her environment. If she receives perceptual information from the latter at all, it is only via other intermediaries over which her intentional awareness seems to lack control: for lack of control is the typical mark of illusory and hallucinatory experience, which are undergone by the perceiver regardless of her ability to direct her perceptual awareness as she wills.⁴⁵ These implications of the argument, as well as other issues,⁴⁶ have led many philosophers to view the argument from illusion as problematic.

⁴⁴ Martin calls this the ‘common kind assumption’: i.e., that ‘whatever fundamental kind of mental event occurs when one is veridically perceiving some scene can occur whether or not one is perceiving’ (2004: 273-274).

⁴⁵ I do not mean to say that a perceiver lacks control over the entire occurrence of an illusory or hallucinatory experience; only that the false phantasms brought by the experience normally come unbidden. How one might be directing one’s awareness *may* still play a causal role in the perception of illusions: for instance, where one seems to see water in the road ahead, such illusion is caused by a trick of the light but also depends on how the perceiver gaze is directed, i.e., the angle of one’s eyes relative to the horizon, etc. Obviously, in some cases there is also voluntary choice and agential control over the steps that one may take (e.g., taking certain drugs) in order to trigger such experiences. There are also cases of illusion, discussed by Armstrong (1968), in which the perceiver may know that things are not what they seem, and yet

The argument from illusion may be of interest to action theorists because it suggests a general causal structure, or what Dancy⁴⁷ calls a ‘constitutive conclusion’, that also characterizes the causal theory of action (1995: 434). The causal theory of action proposes that real action requires the conditions of a physical event, a mere movement, plus some suitable mental event or state, joined together by the right kind of causation. The argument from illusion proposes that true perception requires a mental object or event, plus a suitable physical object or event, with the proper causal relation between the latter and the perceiver’s coming to have its correlative mental object or event in mind.⁴⁸ This analogy to action (let us call it Analogy 1) thus assumes that (a) the common factor of action is a physical event, i.e., a movement, and (b) what distinguishes real action from mere movement is the (analytical) addition of a

the illusion persists: the Müller-Lyer illusion is an example of this, in which one perceives two equal lines as being of unequal length.

⁴⁶ See, for instance, Dancy’s ‘Arguments from Illusion’ (1995).

⁴⁷ Dancy is in accord with the present thesis only up to the point of agreeing that the argument from illusion is just one example of a form or style of argument which appears in other areas of philosophy, besides the metaphysics of perception. Although he reviews the application of the argument from illusion to the case of ‘trying’ (discussed in Chapter 3) and its relation to action, his paper does not address the more fundamental case of mere movement versus action. He is also critical of the kind of reply to causal theories that I will develop in the remainder of the chapter, i.e., disjunctivism.

⁴⁸ Some philosophers, including Grice, P.F. Strawson (1979), and Snowdon, identify a ‘causal theory of perception’ as one requiring the appropriate causal relation between some ‘public object *o*’ and a subject *S*. Snowdon for instance offers what he calls ‘the Effect thesis’: If a subject *S* sees a public object *o* then *o* produces in *S* an L-state namely, ‘a state reportable in a sentence beginning “It looks to *S* as if...” where these words are interpreted phenomenologically (rather than ascribing, say, a tentative judgment by *S*)’ (Snowdon 1980: 176). The L-state produced by the public object *o* has ‘causal immediacy’ in that it is the ‘immediate object’ of *S*’s experience of *o* (see also Putnam 1975 on distal versus immediate causes). Such definitions of veridical perception set the stage for the argument from illusion, whose proponents can use an effect thesis or something similar to it to argue that the production of L-states is possible without the presence of a public or external object *o*. The advantage of such an argument, of course, is that L-states might then be established as another version of sense data.

psychological cause, e.g., mental events or states. In the case of perception, (a*) the common factor is a psychological state that purports to be *of* a physical object, and (b*) what distinguishes real (veridical) perception is the addition of the physical object itself as the cause of the psychological state (the common factor). Let us call the thing that distinguishes real action and real perception in both cases, the ‘distinguishing good cause’ (or DGC).

Both the causal theory of action and the argument from illusion represent a kind of causal view in which a causal intermediary is observed to interpose between the agent or perceiver, and what happens or exists in the public realm beyond the agent or perceiver herself. The causal theory of action thus faces the very same difficulty as is being raised for the indirect theory of perception promoted by the argument from illusion. The argument leads to a picture of the perceiver as disabled from having direct access to her environment, a kind of access that arguably should be hers *as* a perceiver, while the causal theory of action promotes a picture of the agent as disabled with regard to what happens, once the mere movement has been performed, to her body and her physical environment beyond. It is possible to distinguish another analogy at this point (let us call it *Analogy 2*). The indirect theory of perception leads to a difficulty in which the perceiver is cut off or blocked from the physical object that she is supposed to perceive, *even* in cases of veridical perception (where the physical object is truly present). What blocks her access is the psychological intermediary, the common factor (CF), that is in fact caused by the distinguishing good factor in *Analogy 1*. In the case of action, however, the agent is blocked from ‘accessing’ what happens (i.e., causally contributing to what happens) to her body: what blocks her access in this case is the psychological cause of her physical movements.

Whereas *Analogy 2* treats of two ‘good’ cases, a third analogy involving two ‘bad’ cases may also be drawn between action and perception (*Analogy 3*). In the case of action, it may be that the mental events or states occur or are present, and the corresponding

physical movement does not follow. This case is analogous to a perceptual experience in which the psychological state (as of a physical object) is present to the perceiver, but there is no corresponding physical object. In this analogy, the common factor is the same in both cases: the mental state persists regardless of corresponding physical conditions. In the case of perception, this mental state is an effect of a wrong cause, i.e., one that is not the distinguishing good cause. In the case of the action that never occurs, the common factor is an event, a cause that never was, that fails to produce its predicted effects. In both cases, the distinguishing good cause is a physical or external object or event: for perception, it is the corresponding physical object that distinguishes veridical from illusory cases, while for action, the distinguishing good cause is the occurrence of corresponding physical movements that would indicate that one's action succeeded.

The following table outlines the differences between these three analogies:

	Action	Perception
Analogy 1 (Psychological vs. Physical)	CF: physical movement DGC: psychological cause	CF: psychological state DGC: physical object
Analogy 2 (Cause vs. Effect)	DGC (psychological cause) blocks perceiver from CF (physical effect)	CF (psychological effect) blocks perceiver from DGC (physical cause)
Analogy 3 (CF vs. DGC)	CF: psychological cause fails to produce physical effect DGC: physical movement	CF: psychological effect from the wrong physical cause DGC: corresponding physical object

These analogies illustrate, in different ways, problematic conceptual divisions that are present for both cases of action and perception under their causal conceptions, i.e., the causal theory of action and the indirect theory of perception that the argument from illusion promotes. For they each raise a question about the very 'constitutive

conclusion' of the causal theory of action – wherein the assumption of a common factor, a distinguishing good factor, and the attribution of either a mental or a physical event to each of these – arguably provokes some of the problems faced by causalism that I sketched in the earlier part of this chapter.⁴⁹

One major source of this trouble, I believe, can once again be traced to certain Cartesian presuppositions held by the causal approaches so far reviewed. Their shared commitment to a common factor of experience, whether in action or sense perception, entails that the necessary and jointly sufficient constituents of that phenomenon (whether an action or a veridical perceptual event) are of two entirely different types. On the latter view, the mental occurrences share no properties with the corresponding physical occurrences, except in cases where they at least have spatio-temporal properties in common: these are, of course, the veridical cases of seeing something, and the instances of truly realized action. The problem of deviant causal chains, in which the metaphysical gap between agent and action reveals the causal theory's conditions for agency to be insufficient, similarly applies to the conception of perceptual experience suggested by the argument from illusion. Suppose that Descartes's evil demon were to produce a hallucination of a poinsettia plant at the very moment that I turn my attention to gaze at a real poinsettia plant in my immediate surroundings. Suppose further that the evil demon always acts such that whenever I deliberately turn my attention to an external object, my movement to look causes him to immediately produce an exact mental replica of whatever it is that I was about to veridically see. One cannot possibly claim that my perception, supposing that what I see is the hallucination, and nothing else, is a veridical one. The sufficient conditions for veridical perception are in place: I have a mental object in my mind, which corresponds to a physical object that really exists in my environment. As in the

⁴⁹ One problem that may be thought to arise from Analogy 1, for instance, is the problem of deviant causal chains. In my next chapter, I offer some detail on how this might come about.

case of action causalism, here there is also reason to doubt that what the causal theory cites as sufficient conditions for true action and perception are really sufficient. For the case of perception as well, the causal theory presupposes a metaphysical gap between constituent types, a gap in which deviant causal chains might be inserted.⁵⁰

Other philosophers have identified the particular epistemological features of this Cartesian picture⁵¹, of the causal approach to both action and perception, as worrisome. Byrne and Logue characterize the argument from illusion as producing ‘the Cartesian view’ of perception, according to which ‘in both the good case and the bad case you are having exactly the same kind of perceptual experience’ (2008: 10). Haddock and Macpherson also recognize the causal theorist’s notion of a common factor of experience, but identify it further as an extension of the background assumption that the constituents of experience are of entirely different types. They write:

The Cartesian picture of the mind understands the inner world as constitutively independent of anything outer: it is possible that the outer world differs radically, and the inner world remains exactly as it is. There are various ways of understanding the notion of an inner world. We

⁵⁰ This hypothetical involving a hallucination of a poinsettia plant can be seen as an iteration on Analogy 3, which suggests that in both cases of action and perception, the normal common factors are in place (i.e., I have a vision of a poinsettia plant where my perception should be of one such object), but the physical cause is flawed, even where there is actually the correct corresponding object present (i.e., Descartes’s evil demon causes my awareness to be of his hallucination rather than of the true physical object).

⁵¹ By emphasizing the epistemological features of his view, I mean that the trouble here is not necessarily one that relates to Cartesian metaphysics of mind-matter dualism, or even property dualism that might be thought more acceptable. A more basic epistemological point for Descartes arises early in the *Meditations*, before Descartes has even concluded that there is such a thing as matter as essentially different from mind (*Med.* VI): he muses that whether or not he is dreaming of sitting by the fire, his experience would be exactly the same.

might understand it as the idea of (amongst other things) the subject's experiences (what Snowden calls their L-states). We might understand it as (simply) the idea of a subject's experientially acquired reasons – those reasons that a subject enjoys when it looks to him as if things are a certain way. And we might also understand it as the idea of the phenomenology of experience, perhaps for the reasons McDowell (1986) suggests, namely, that it is the idea of what is within the scope of the subject's capacities for infallible knowledge (2011: 22).

However the 'inner world' of a perceiver is understood (the next section ventures more in depth on this topic), Haddock and Macpherson's main point underlines a difficulty with the causal theory. The constitutive independence of the inner world from its outer environs renders it impossible that the mental might at least be necessarily correlative with the physical; in fact, 'the inner world remains exactly as it is' while its outer counterpart endures variable change, or perhaps the perceiver's 'outer world' remains constant while the 'inner world' differs radically. Dancy remarks that 'the idea that arguments from illusion stem from defective Cartesian presumptions about the independence of the mental...is a ground for treating such arguments with suspicion' (1996: 438). In this chapter, I have suggested that the suspicion is warranted, especially as such 'independence' is understood to apply to our epistemological considerations.

The argument for sense data from illusion, like the causal theory of action, promotes a variety of different ways to understand the causal relation between mental occurrences and physical facts or objects. Haddock and Macpherson, in the passage above, mention at least three possible ways of understanding an 'inner world' of a perceiver, and each of them merit a corresponding reply. All of these replies fall under the heading of an approach known in the theory of perception as 'disjunctivism'.

3.2 Action disjunctivism

Disjunctivism in the philosophy of perception represents a relatively recent response to the argument from illusion. Some philosophers believe that disjunctivism amounts to no more than a bald denial of the indirect causal theory of perception (Dancy 1995). Others defend the approach as a substantial thesis and compelling alternative to the causal theory. Byrne and Logue explain: ‘According to disjunctivism, the good case and the (hallucinatory) bad case share no mental core’ (2008: 10). Although this statement sounds like a bald denial of the causal theory of perception, whose commitment to a shared factor of experience, a ‘mental core’, was made evident in the previous section, there is more to the disjunctivist’s claim. In this final section, I will consider its usefulness for offering a model for a substantial alternative to the causal theory of action, by considering its strengths in the face of certain objections. First among them is the claim that disjunctivism is an idea that lacks unity among its proponents. All disjunctivist approaches reject the causal theory’s idea of a shared common element, but disjunctivists often disagree over, e.g., the implications of making sharper contrasts between veridical and illusory cases, and the extent to which these cases diverge. I begin by summarizing three very influential disjunctivist approaches, covering the three ways cited by Haddock and Macpherson concerning possible conceptions of a perceiver’s ‘inner world’, with the overall aim of attending to certain important differences among them.⁵²

First, Martin advocates a variety of disjunctivism that he calls ‘naïve realism’. He defines disjunctivism as requiring ‘us to view the transparency and immediacy of perceptual experience as involving actual relations between the subject and the objects of perception and their features. In just the case of veridical perception, the experience is a matter of certain objects being presented as just so’ (Martin 2002: 402).

⁵² The summary that follows is a synopsis of Haddock and Macpherson’s very incisive Introduction, which offers an explication of the three major strands of disjunctivist thought (2011).

In cases of illusory experience, the lack of a true object of sight entails that there is no such presentation (of the object ‘as just so’), and no actual relations between the subject and the objects of perception (and their perceptible features). In light of these deeply metaphysical contrasts between the veridical and the illusory cases, Martin argues that the phenomenal characters attributed to these respective experiences belong to different ‘fundamental kinds’ (Martin 2002: 404; 2004: 43, 54, 60).⁵³ Haddock and Macpherson remark, ‘This response to the argument from hallucination denies the highest common factor conception,⁵⁴ which claims that what explains the phenomenal character of the experiences is the same in all cases. And it does so by denying that the experiences have the same constituents and are of the same fundamental kinds’ (2011: 16-17).

Snowdon’s disjunctivist approach differs from Martin’s formulation (in both their earlier and later versions) in a few respects. First, he does not deny the causal theorist’s premise of phenomenological indiscriminability. Second, he attempts to

⁵³ Martin has changed his view, however, on the question whether experiences of veridical perception are truly indiscriminable from instances of illusion or hallucination. He clarifies the notion of indiscriminability as perhaps holding in the sense of a perceiving agent’s ‘introspection’, or being based on how an experience ‘feels’ when the perceiver pauses self-consciously to consider its phenomenal qualities. In the previous section, consideration of the ‘Cartesian picture of the mind’ suggested that phenomenological indiscriminability constituted an indispensable assumption for the argument from illusion. In later writings, however, Martin commits to the idea that the phenomenal quality of experiences in the veridical case need not be the same as the phenomenal quality of the hallucinatory experience (Gendler and Hawthorne 2006: 366-372). It seems right to conclude that Martin’s disjunctivist argument hinges on the notion of ‘fundamental kinds’, whose distinction depends on phenomenological, introspective discriminability, as well as the real presentation of external objects ‘as just so’.

⁵⁴ This term, ‘highest common factor conception’, can be taken to be equivalent to what I have been calling the common factor of experience, an important feature of causal theories of action and perception (§3.1). The highest common factor claim is integral to the intentionalist theory of perception, which understands perception as representation of an object, such that the phenomenal character of experience is determined by mind-dependent features of that representation.

give an argument that does not refer to the causal relations between external object and perceiver. Instead, he employs a distinction between the internal, mental states of the perceiver, states which he calls L-states. Snowdon explains, ‘According to the disjunctive theory, the explanation of *S*’s failure to see *o* is not that the L-state is not caused by the object seen but (simply) that *S* is in the *wrong kind* of L-state; and, in the same way, the explanation of *S*’s success in seeing *o* would be (simply) that *S* is in the right kind of L-state. The theory offers the following account of what it is to be in an L-state: it is either to be in an L-state that is ‘intrinsically independent of surrounding objects’ or to be in an L-state that intrinsically ‘involves the surrounding objects’ (Snowdon 1980: 186).⁵⁵ Thus, if the perceiver is in an L state of the latter kind, then she does see real objects (and the objects she sees are the very objects that being in such a state ‘involves’); and if she is in an L-state of the former kind then she does not. Snowdon distills his theory by means of the following disjunction: ‘[D] It looks to *S* as if there is an *F*; (there is something which looks to *S* to be an *F*) ∨ (it is to *S* as if there is something which looks to *S* to be an *F*)’ (Snowdon 1980: 185).

McDowell offers a very different variety of disjunctivism from those of both Snowdon and Martin. His approach is known as ‘epistemological disjunctivism’. In contrast with the so-called metaphysical or ‘experiential disjunctivism’⁵⁶ proposed by Snowdon and Martin, which addresses the relation between the external object and its perceiver, McDowell’s approach focuses on the different kinds of reasons a perceiving agent might give for believing that *p*, based on appearances, in either the case of true perception or the case of illusion. Presuming phenomenological indiscriminability,

⁵⁵ Although it is not my aim to evaluate the success of each of these disjunctivist accounts, it is not clear to me that Snowdon’s account succeeds on this point. If his aim is to give an argument that does not refer to the causal relations between external object and perceiver, and yet the difference between L-states is defined in terms of their either being ‘intrinsically independent of’ or ‘intrinsically involving’ surrounding external objects, then it looks as if some causal description concerning such objects and perceiver is in order.

⁵⁶ Haddock and Macpherson (2011: 16).

McDowell argues that, despite the fact that the true case is ‘just like’ the false case (whether phenomenologically or through introspection), there exists a fundamental difference between the experiences. In the true case, the perceiver is able to cite reasons that are simply unavailable to the perceiving agent who experiences an illusion or hallucination. As for Martin, however, a key premise in McDowell’s arguments is the idea of a fundamental difference.⁵⁷ However, both Martin and Snowdon entirely abandon McDowell’s explanations concerning the kinds of reasons that our experience supplies to us as perceiving agents. The difference between the experiential and epistemological varieties can be made even sharper: Byrne and Logue, for instance, observe that ‘epistemological disjunctivism is quite compatible with the denial of metaphysical disjunctivism’, the varieties proposed by Martin and Snowdon (2008: 67).

As may be expected, the disjunctivist approach, understood generally as a rejection of the causal theory’s idea of a shared common element, has its critics. Perhaps the most fundamental challenge to the disjunctivist approach is with regards to its status as a truly substantive alternative to the causal theory of perception. Critics claim, for example, that disjunctivism is merely a bald denial of the indirect causal theory of perception, nothing more than a quick, facile rejection of the argument from illusion. Dancy, for instance, offers the following critique:

[Disjunctivism] amounts to the assertion that there are broadly two types of experience, which we can call manifesting and non-manifestings, and that some of the latter are indistinguishable (to their owners, at the relevant time, etc.) from the former. We are told virtually nothing about the natures of these different types of experiences. What sort of answer, then, does disjunctivism offer to the argument from illusion? In what way does it go at all beyond...my description of what is required if one

⁵⁷ See Martin’s ‘The Transparency of Experience’ in *Mind and Language*, vol. 17 (2002).

wants to reject the argument from illusion? I do not think it goes any further at all. It just repackages the idea that those who reject the argument from illusion in the perceptual case are bound to say that there can be two states of mind which are indistinguishable to their owner though neither is an ingredient in the other. It makes no move *beyond* this at all. The disjunctive conception, then, does not amount to more than a structured way of expressing opposition to the argument from illusion (1995: 435).

If Dancy is right, then what was thought to serve as a substantive alternative to the Cartesian view is merely the stylized ‘repackaging’ of a superficial negation of the causal theory. Disjunctivism does nothing more than merely adopt a position contrary to the common factor of experience premise: disjunctivism simply rejects that notion, and splits perceptual experiences into two ‘fundamentally’ different types, without any shared constituents between them, besides phenomenological similarity or indistinguishability. Furthermore, what grounds the ‘fundamental’ difference between them is not sufficiently explained, complains Dancy. In the absence of a substantive explanation of this difference, disjunctivism apparently merely offers a re-statement of the problem of perception, leaving a real solution in the offing.

Disjunctivism faces another related challenge apropos its proposed status as a substantive alternative theory to the argument from illusion. Given the significant disagreements between disjunctivists themselves, as exhibited by the foregoing discussion differentiating Martin, Snowdon, and McDowell, someone may doubt that there is such a thing as *the* disjunctivist approach. Some disjunctivists adopt phenomenological indiscriminability between veridical and illusory cases as an essential premise of their approach, while others argue that such indiscriminability is necessarily false. Philosophers have noted that epistemological disjunctivism may even be compatible with the rejection of metaphysical disjunctivism (Byrne and Logue

2008). These and other disagreements may lead one to doubt that there exist any common links at all between even these three varieties of disjunctivism alone.

There may be an obvious solution that quells both of these preliminary objections.

Many philosophers, including Dancy, presume that the different varieties of disjunctivist approaches should be analysed as standing harmoniously opposed to ‘the Cartesian picture of mind’, as the latter is exemplified in perceptual theories. As discussed in the previous section, the indirect causal theory of perception adopts the argumentative strategy of citing an event-type, a mental object or occurrence, as one of the sufficient conditions for perception, whether veridical or illusory. The disjunctivist approach rejects the common factor of experience premise, as an explanation for why the veridical cases often seem identical, or at least similar, to the cases of false perception. Thus far, this characterization appears to remain subject to Dancy’s critique of disjunctivism. But disjunctivism goes further than the facile conception he imagines. In its rejection of a common factor of experience, disjunctivism also resists a fundamental premise of the causal theory, i.e., the Cartesian divide between the mental and physical constituents of perception, or what Haddock and Macpherson call the ‘inner’ and ‘outer’ worlds (2011). Disjunctivism, in contrast, offers the alternative of understanding these domains as sometimes interconnected, or somehow ‘permeable’ to one another. Thus the physical or external objects of one’s perceptual awareness, although not literally existing within the domain of the perceiver’s mind, are nevertheless truly constitutive of the perceiver’s veridical experiences. The same, of course, cannot be said of the cases in which the ‘immediate’ objects of my perception are illusory or hallucinatory. For these objects are sensory errors, a mistake of the body precisely because the inner and outer domains are not appropriately interconnected or made permeable to one another. Haddock and Macpherson offer these comments on the mental and physical permeability as the fundamental basis for veridical experience, across all three varieties of disjunctivism reviewed here:

[W]e can see that Snowdon's metaphysical disjunctivism opposes the Cartesian picture once the idea of the inner world is understood to include a subject's experiences, for it understands some of these as constituted by outer objects and facts. McDowell's epistemological disjunctivism also opposes it, once the idea of the inner world is understood as the idea of the subject's experientially acquired reasons, for this variety of disjunctivism insists that, in some cases, these reasons can be to the effect that she sees that *p*. And Martin's metaphysical disjunctivism also opposes it in the experiential case, once the idea of the inner world is understood as the domain of experiential phenomenology, for it insists that this can sometimes be explained (and even constituted) by outer objects and facts (Haddock and Macpherson 2011: 22).

By rejecting the metaphysically dichotomous view of perceptual experience, the varieties of disjunctivism are unified in their attempt to capture a picture of veridical experience as requiring both physical and mental elements at once, involving both the phenomenological world of felt objects and the inner world in which such objects are also included, although not physically present.⁵⁸ Despite the cited differences among the three main varieties of disjunctivism, Haddock and Macpherson therefore conclude: '[T]he mark of disjunctivism, in all of its varieties, is a conception of the inner and the outer as suffused' (2011: 22). Disjunctivism, on their account, overcomes the argument from illusion by asserting that false perceptual experiences fundamentally lack this quality of inner and outer 'suffusion'. That there exists a 'fundamental difference' between veridical experience and illusory experience, constitutes a general notion that unifies the aims of the disjunctivist approaches reviewed.

⁵⁸ Specifically *how* they are present in the 'inner world' is the subject of much debate. Aquinas, for instance, argued that the mode of existence was *esse intentionale*, or intentional being, rather than natural or actual existence; Brentano later took up this theme (1874).

The concept of mental and physical ‘suffusion’ is apparently employed as a metaphor for the concept of veridical experience that the disjunctivist wishes to illustrate: the metaphor represents a relation of direct access between the perceiver and a real object in her immediate environment. The perceiver is not alienated from objects beyond her mind, on this picture; rather, her experience of them is one of complete involvement. It is obvious that the concept of ‘suffusion’ will require careful philosophical development. Nevertheless, it can already be anticipated that what Haddock and Macpherson call the ‘suffusion’ between the mental and the physical constituents of perceptual experience⁵⁹ arguably holds significant implications for the problem of causal exclusion raised earlier in this chapter. If the core thesis of disjunctivism is true, then it is possible to understand the ‘inner’ and ‘outer’ realms of experience as inter-permeable, as open to one another; and there is also good reason to believe that the concept of metaphysical suffusion thus far understood for veridical perception might be applied to analysis of action as well. Disjunctivism, as an account of action distinguished from mere movement, may be developed as a substantive alternative to the causal theory of action.

The previous section explored the possibilities in which the causal theory of action may be understood in a way analogous to the understanding of perception that gives rise to the argument from illusion, i.e., the indirect causal theory of perception that the argument promotes. Like the argument from illusion, the disjunctivist approach to perceptual experience may be thought to represent just one form of an argument that also appears in other areas of philosophy.⁶⁰ Its application to action theory is a prime example. A disjunctivist approach to action, then, not only rejects the causalist notion of a common factor of experience, whether action or mere movement. Action disjunctivism aims primarily to achieve a metaphysical ‘suffusion’ between mental and

⁵⁹ It is possible to see in this concept the potential for development into what Hanna and Maiese call mental and physical ‘property fusion’. The possibility of metaphysical fusion between the mental and the physical is the subject of the third chapter.

⁶⁰ I do not review the different applications here, but Dancy offers an overview (1995).

physical occurrences, of the kind proposed by Haddock and Macpherson. Its application to action is based on comparisons from its core thesis for perception:

Something similar can be said for disjunctivism about bodily movement. We might understand the idea of the outer as consisting at least of the subject's bodily movements. We might understand it as (simply) the idea of those of a subject's reasons that concern the behavior of others reasons that specifically concern how things are with the bodies of others. Metaphysical disjunctivism about the surface sees some bodily movements as partly constituted by inner states. And its epistemological variant sees some of the reasons that concern behaviour as also concerning the state of the other's inner world (as when subjects see that the other's behaviour is expressing the fact that they are in pain). We might think of disjunctivism about bodily movements as an inverted, but structurally analogous, version of disjunctivism about experience, because it asserts, not the suffusion of the inner by the outer, but that of the outer by the inner. Put together, both kinds of disjunctivism yield a conception of the inner and the outer as each suffusing the other (2011: 22).

Disjunctivism was earlier considered in light of the criticism that it fails as a substantive theory, merely 'amount[ing] to the assertion that there are broadly two types of experience, which we can call manifestings and non-manifestings' (Dancy 1995: 435). Even if Haddock and Macpherson's description of action disjunctivism falls short of a complete explanation, their proposal, if it is rigorously pursued, holds enough promise for one to believe that the disjunctivist approach in general amounts to much more than a brute juxtaposition of two fundamentally different kinds of events (generally conceived), or perhaps a quick and facile assertion that there exists no common element whatsoever between them. Disjunctivism, whether applied to the

analysis of perception or action, hinges on the intriguing picture of a single realm in which agential phenomena, both mental and physical, transpire. The argument from illusion is motivated, obviously, by the fact that not all perceptual experiences are veridical, although many of these experiences may *appear* to be veridical. The solution offered by disjunctivism expresses a metaphysical claim about the way veridical experiences are distinguished from perceiving things falsely, regardless of the phenomenal discriminability or lack thereof between the cases. Disjunctivism proposes that we look instead for a constitutive relation as our starting point, and try to understand a conception of a perceiving agent's mental processes as being partly constituted by physical objects or facts. In veridical cases, one's phenomenological experience of seeing an object is therefore constituted by the way things really are in the surrounding environment, in such a way that one's perception of the poinsettia is really of the poinsettia itself.

Similarly, the problem of causal exclusion (and mental causation in general) is partially motivated by the fact that not all physical movements are actions, in the specific sense of being voluntary, intentional, or rational, although a mere movement and an action may share certain identical features and descriptions. The solution offered by disjunctivism distinguishes action from mere movement by promoting a metaphysical picture of inner, mental events as constitutive of certain outer, physical events. An epistemological variant involving reasons might in fact be derived from the metaphysical or experiential version, since it is possible to see how the 'suffusion' of inner and outer realms provides the basis for an agent to give reasons for action of one kind over another: 'suffusion' in agency may potentially be understood as occurring wherever an agent is justified in assuming her external situation, including the fact of her own body, and where agential guidance or control over that body is realized. Where such suffusion is lacking, and the movement is involuntary, an agent's

explanation of her movements will be of a very different kind⁶ⁱ: for instance, one may attribute the cause to a case of nerves or a muscular disorder. The mental conditions and ‘inner’ features of agents have a varied form of existence: they may take the form of a person’s history of deliberation, if one exists, or they may be understood as psychological states of mind or mental attitudes possessed by the agent; in general they may be said to constitute the constant and active attentive hum of all motile and sensate animals. However we characterize the minds of agents, such mental conditions and features are essentially constitutive of actions, in such a way that the mind may be said to inform the very physical existence of an agent’s intentional movements. The very concept of action somehow relies on the ‘suffusion’ of body and mind. It follows from this idea that disjunctivism, as applied to action theory, rejects the idea that action and mere movement possess a common constituent, the causalist notion of a mere movement. Involuntary, non-voluntary, and unintentional events are of wholly disparate kinds from true action, because these lack altogether the necessary ‘suffusion’ of mind and matter. It is arguably easier to see the disparity between action and mere movement than between the disjunctive elements in the perceptual case: for usually it is clear to us whether reasons for action are present or absent, and whether a body’s movements are governed or directed by the agent or not regardless of deliberation or mental states.

Much more than a bald or facile rejection of the argument from illusion, disjunctivist approaches to action therefore open an important avenue for overcoming a tendency in the philosophy of action to presume that the causal theory of action is the only plausible explanation of action. The notion of a common factor of experience, a significant feature of causalist accounts in general, is by itself vulnerable to the

⁶ⁱ Some may be interested to recall that Anscombe makes this point; to an extent, one might be able to describe her account as ‘disjunctivist’ as well.

objection that involuntary behavior⁶² — such as breathing, blinking, and curling one’s toes while walking — are not events to which one simply adds a mental cause to produce action. We can certainly do all of these things voluntarily. But to do so is to engage in a different kind of behavior entirely, with a very different material occurrence to characterize it. Action causalism, for this reason, is unsatisfying. Among the various other objections I have reviewed in this chapter, one may add that it is overly simplistic and thus, for this reason also, phenomenologically deficient in its characterization of voluntary agency. For the phenomenological qualities of action versus involuntary movement, it must be admitted, are quite different. Although this difference may be less obvious in the case of perceptual illusion or hallucination, it is much more difficult to doubt whether one’s action is one’s own, even if it were to occur in a dream: for even in dreams, Descartes would have to admit, one’s voluntary movement is qualitatively different from those that are involuntary. In action, our sense of volition and control over ourselves is retained, whether our experience of agency is illusory or not.

In this chapter, I have suggested that disjunctivism plausibly opposes the Cartesian schema of inner worlds untouched by outer worlds. It does so by positing some kind of metaphysical picture of ‘suffusion’ between the inner and outer worlds of a perceiving agent, with the result that the concept of action cannot be easily divided into two constituents that are opposed in kind. Action disjunctivism also accords better, I believe, with our first-person experience of

⁶² Involuntary behavior is characteristically distinct from nonvoluntary behavior in that the former can be made voluntary with intervention. Nonvoluntary acts are of the kind that, normally, attempts to intervene cannot change their patterns of movement. There are extreme exceptions to this description: some individuals seem able to alter the rate of their beating heart, or even to move their own digestive tract upon mental command.

voluntary and intentional⁶³ movements as being phenomenologically unified across mental and physical realms, with the result that it is almost never necessary to distinguish⁶⁴ them.⁶⁵ For most of what we do, the spontaneous performance is one that combines the cognitive dimension of my action (e.g., that I know I am *A*-ing) and its sensitive or material aspects (e.g., that I sense my body moving so that I do *A*). The phenomenology of such mental and physical suffusion is obviously absent in cases of involuntary or unintentional movements, in which there may perhaps be a sensitive element unaccompanied by a cognitive aspect: e.g., an agent may suddenly realize that he is breathing rapidly out of fear or anger. In its rejection of the notion that physical objects are wholly external to perceiving agents, and thereby inaccessible to one's internal experience, disjunctivist approaches to action provide a promising thesis for distinguishing true action from mere physical movement.

Disjunctivism as applied to action, as presented in the present chapter, still faces many unanswered questions.⁶⁶ The first difficulty, as I have already mentioned,

⁶³ I do not mean to use the terms 'voluntary' and 'intentional' as conceptually equivalent. I recognize and discuss the difference between them in my second chapter.

⁶⁴ Thus far I have assumed that any phenomenological distinctions to be made are made by us as agents ourselves, rather than as external spectators of other agents' movements. Whether or not we as external spectators likewise have the ability of distinguishing whether an animal's movements are the consequence of a reflex or constitute an action is a different question, one that I am not at leisure to discuss in this thesis.

⁶⁵ Trying to accomplish something very difficult, on the other hand, may produce frequent reminders of the divide between what I wish to do and my actually doing it. Playing the Paganini Caprices for violin is a suitable example for most people.

⁶⁶ Some philosophers may also object that the disjunctivist view resembles the original Cartesian argument about rational action as proceeding from mind, not nature. I believe that arguments can be provided to demonstrate that this impression is false. For while Descartes's thesis focused on the causation of mind moving matter, disjunctivism about action is a thesis that focuses rather on fundamental differences between kinds of matter (that is, physical

concerns a precise explication of the metaphor of ‘suffusion’ between inner and outer realms. A related question may be raised concerning the possibility of ‘property fusion’, an explanatory or metaphysical relation, between mental and physical properties. The direct theory of perception, a close relative of disjunctivism about perception, also continues to be beleaguered by worries relating to the question how an external object can come to be ‘inside’, or in direct contact with, the perceiver’s mind without literally existing in her mind. This particular issue dates back to an ancient debate over *esse intentionale*, the claim that an external object may have ‘intentional existence’, contrasted with material existence, within the perceiver’s mind. Action disjunctivism, thus far characterized, also leaves undecided any precise conceptions concerning the nature of mind and its ‘contents’ which, suffused with the material occurrences involving the body, describe action in a non-causalist way. The next chapters are devoted to explicating the idea of mental and physical ‘suffusion’, known in other contexts as psychophysicality, with action being the prime example.

movement) for which mindedness is significant. Two important differences between disjunctivist approaches to action and the Cartesian schema are especially notable. First, although Descartes recognized the need for establishing a unity in action between mind and matter, he was unable to achieve it (or indeed to offer a satisfying explanation for it) due to his inability to relinquish the division between *res cogitans* and *res extensa*. On his metaphysical picture of mind bifurcated from body, the dynamic suffusion of inner and outer realms was impossible (or highly implausible): the explanation for any interaction between them speculatively located the mind-body nexus within the pineal gland. Second, Descartes’s view of agency hinged on the idea that action was distinguished by the triumph of human rationality over animal nature. Nothing in disjunctivism about action as such requires one to hold this thesis; the approach that I adopt in this thesis certainly rejects it. Instead, my disjunctivist approach to action embraces a view of mind as biological, as a feature of the natural world. In this latter sense there is a clear Aristotelian thread that runs throughout my thesis. The idea of inner-outer suffusion, or psychophysicality, as a concept inspired by Aristotle, constitutes the subject of the succeeding chapter. Some recent philosophers working in a similar vein, notably O’Shaughnessy, even attempt to capitalize on Descartes’s attempt to solve mental causation by arguing for a concept of ‘action as the pineal gland’. I review the modern version of this attempt in the third chapter.

3.3 Transitional remarks

Before moving on, I would like to make an important preliminary note concerning the logical structure of the remainder of the thesis. The succeeding chapter introduces interpretations of Aristotle's thought on the question of what distinguishes an action from a mere movement. This apparently historical transition is warranted for several reasons. First, contemporary action theorists claim Aristotle⁶⁷ as exerting the greatest influence over the formulation of their core questions. These include, but are not limited to: 'What is action as distinguished from mere movement?', 'How is action produced; what are its causes?', and 'What are reasons for acting?'. As I explain in Chapter 2, proponents of the causal theory of action in fact claim Aristotle as the principal progenitor of modern causalism. Even if this latter claim could be successfully defended (and I will argue that it cannot), by that fact alone there would be sufficient interest for investigating where Aristotle stood on the matter. The current chapter concludes with the tentative suggestion that the disjunctivist's metaphor of 'suffusing' the mental and the physical realms might be clarified to overcome a Cartesian picture of mind and matter. This Cartesian picture, I have suggested, is a critical premise in the account provided by the causal theory of action. Aristotle, I will argue, is an opponent rather than an advocate for the causal theorist. For he defended an account of mind and matter so unlike the Cartesian picture that the two are rightly understood as conceptual enemies.

⁶⁷ Consider, for instance, Anscombe's *Intention*, which Davidson famously describes as 'the most important treatment of action since Aristotle'. Besides Wittgenstein, comments Stoutland, 'the other great influence on Anscombe's work is Aristotle. Very many of the key ideas in *Intention* use Aristotelian terminology and assume Aristotle's discussion. I list just some of them: practical reason, the practical syllogism, practical knowledge, the centrality of ends in action, the concept of desire and wanting, human agents as animals, the notion of form, and the idea of something being what it is by its nature. *Intention* is permeated with Aristotelianism' (2011: 6).

Aristotle, I will endeavor to show, proposed an account of psychophysicality manifest in action and perception. Even if his account falls short of constituting a true theory of action, since he nowhere intended to address the questions that concern today's action theorists, Aristotle's understanding of action as psychophysical is of immediate interest because it promises a further explanation of the disjunctivist metaphor, i.e., of 'suffusion' between the mental and the physical. His concept of action as a species of change, as it will turn out, provides the basis for construing him as an originator of the disjunctivist's core idea of distinguishing good and bad cases by way of fundamentally different kinds. As I will argue in the next chapter, Aristotle's thought on action deserves careful treatment not least of all because he may be understood (however anachronistically in our terms) as an early proponent of a disjunctivist approach to the concept of action.

Chapter 2: Aristotle as Philosopher of Action

I engage with possible diverse interpretations of Aristotle as a philosopher of action, after acknowledging some initial doubts over whether such an undertaking is even possible. My inquiry is divided into two general parts. First, I examine the case for thinking that Aristotle was a kind of action causalist. Causal theorists, beginning with Donald Davidson, are often wont to claim Aristotle as the progenitor of their own views, and I examine the sources for substantiating this claim. Having found the case for making it unjustifiable, I examine a moderately contrasting position that recognizes the demerits of the causal theory while defending the idea that Aristotle accepted some very generalized causalist theses. This approach focuses on a concept of action as Aristotelian change, while also suggesting an interpretation of Aristotle as the inspiration for modern disjunctivist approaches to action. I consider several significant deficiencies of the latter interpretation of Aristotle, with the conclusion that its problems outweigh its advantages. Thus the second general part of my inquiry addresses the more nuanced, but also somewhat controversial view of action as a psychophysical or hylomorphic process. The overall aim of this chapter is to consider the extent to which Aristotle's concept of action is able to vindicate the disjunctivist approach to action, which previously was deemed lacking in substance (§3.2 in Chapter 1).

1.1 The possibility of an Aristotelian theory of action

Is it possible to regard Aristotle as an interlocutor among 20th- and 21st-century philosophers of action? In particular, is it possible to interpret Aristotle as engaging with and answering a core question in contemporary philosophy of action, sometimes referred to simply as 'the problem of action', and illustrated by the question, 'What is

the difference between the rising of my arm and my raising it'? The concern of this chapter is to evaluate different responses to this question that have been defended, in one way or another, as distinctively Aristotelian. Presuming that there possibly *is* a sense in which Aristotle's approach to action resonates at all with contemporary philosophers of action and psychology, the aim will be to uncover common points of interest, while also noting where the accounts may diverge. For starters, there are obvious reasons to believe that Aristotle's aims and concerns were very different from those of contemporary philosophers, and that any claim to the contrary is simply anachronistic. It can indeed be demonstrated that Aristotle presumed altogether disparate starting points from those of contemporary theories, and that he was therefore interested in altogether different questions. On the one hand, then, Aristotle might have rejected, in their entirety, most contemporary schemes that describe human agency, making it impossible for us to regard him as a philosopher of action in the modern sense. Certainly, his work cannot be construed as engaging directly with any particular set of contemporary views. But there is also evidence that Aristotle might have affirmed various features found in contemporary philosophy of action. In addition, it may be possible to develop a solution to the problem of action on Aristotle's behalf, as well as what might be identified as a full-fledged Aristotelian theory of action. My aim in this chapter is to evaluate competing interpretations that seek to accomplish precisely these latter two aims.

One interpretive account to which I will give considerable attention is that offered by Coope, in 'Aristotle on Action' (2007). Coope interprets Aristotle as offering an account of action as change: a notion of action that, on Coope's interpretation, holds some elements in common with causalist theories such as Donald Davidson's, while also resonating substantially with critics of causalist approaches, such as disjunctivists about action.⁶⁸ Because Aristotle appears to agree with certain key elements found in

⁶⁸ I present the alternative offered by disjunctivists in my first chapter, 'Disjunctivism in Action'. In general, my thesis treats disjunctivist theorists as proposing a

causalism and diverging from others, it is claimed that his approach succeeds in entirely circumventing the problem of action as well as the issue of causal exclusion. According to Coope, Aristotle offers a particular conception of agency that eludes most philosophers: with their contemporary post-Cartesian and neo-Platonic commitments, they simply cannot conceive of his concept.

It is expected that some philosophers will be initially skeptical of the characterization of Aristotle as having a theory of action at all. For one, the very fact that Aristotle's views appear to bypass contemporary difficulties in mental causation and agency might be taken as evidence that the concerns with which he dealt now have little relevance for contemporary theories. He never explicitly treats the problem of action or mental causation, and there is little or no evidence suggesting that such questions could ever even arise given his theoretical framework. When he speaks of physical movement as a species of change, Aristotle's explanations involve certain metaphysical notions, such as potentiality and actuality, which at best appear opaque to modern theorists. So there seem to be good reasons to reject a presumption that the very enterprise with which Coope and other commentators are concerned should even be undertaken. For this kind of skeptic, any attempt to do so regardless will ultimately fall short of its aim to establish Aristotle as holding a view that is competitive or comparable with other theories, because this simply cannot be done.

In his own writings, Aristotle never explicitly discusses the concept of intention or intentional movements. The term Aristotle uses is *hekousion*. In translating it into modern English, some commentators equivocate between the 'intentional' and the

paradigmatic critique of the causal theory of action; there may well be many other critical approaches, however, which I neglect entirely and yet are worthy of close examination. Anscombe's treatment of the concept of action (1957) is an example of a critical alternative to the causal theory of action. Although one would be hard pressed to claim that her approach is 'disjunctivist', I believe that one could make the case that her thought *tends* in the latter direction. Whether or not this idea can be justified is clearly beyond the current scope of the thesis.

‘voluntary’.⁶⁹ But these concepts cannot be equivalent, especially given that something can be done intentionally but not voluntarily, even by Aristotle’s lights⁷⁰, and since the term ‘voluntary’ is also ascribed to movements (and results in general) that are unintended although foreseen. My convention, therefore, will be to treat the term as properly translating into the ‘voluntary’. If this analysis is right, then we may doubt all the more that it is possible to even attempt to uncover a definitive theory of action for which the concepts of intention and ‘intentional action’⁷¹ are of special interest that can be ascribed to Aristotle.

⁶⁹ Charles for instance wrongly equates the intentional with the voluntary. Charles translates *hekousion* and cognates by ‘intentional’, etc. (1984: 61-62, 256-61). Also see Charles’s essay ‘The Eudemian Ethics on the Voluntary’ for a similar canvass on these views, in *The Eudemian Ethics on the Voluntary, Friendship, and Luck*, ed. Fiona Leigh (2012). Broadie comments: ‘What is in common, so far as I can see, to all the items which Aristotle terms “*hekousia*”, is that in one way or another the agent says “Yes” to their being or becoming, whether through affirmation, compliance, or failure to say “No”... Charles (62) suggests that what one brings about knowingly (hence acceptingly) but not intentionally should be regarded as “intentional in a *derived sense*”. This artificially coined sense of “intentional” seems designed to encourage us to translate Aristotle’s “*hekousion*” as “intentional” in all its occurrences and to persuade us by this usage that there are correspondingly primary and derivative senses of “*hekousion*” in Aristotle. But Aristotle seems not to differentiate. That is to say, while he does recognise that *hekousiai* actions are primary, whereas the states etc. knowingly produced by them are *hekousia* only because the actions are, he shows little sign of accepting any such proposition as that the latter are *hekousia* in a different or weaker *sense* than the former.’ Broadie continues, ‘That the concepts are different is clear from the facts (1) that Aristotle regards as *hekousia* acts done from culpable negligence; (2) that he regards as *hekousia* all foreseeable consequences of what we do *hekotes*.... See Heinaman [1] for a defence of the traditional translation, “voluntary” (1991: 132, fn. 10).

⁷⁰ In *Nic. Ethics* III, Aristotle offers the case of a man who is coerced by a tyrant to do what the latter wishes on pain of death. The man’s subsequent actions are intentional, but his voluntariness is diminished (if not altogether absent) given the man’s state of reluctance or fear, which lessens the extent of his culpability, and thus also his voluntariness.

⁷¹ As in other chapters, my convention will be to treat the term ‘intentional action’ as redundant of the concept of action itself – although I employ the redundancy here in a style familiar to 20th and 21st century action theorists.

On the other hand, it is widely claimed that there exists a variety of ways, both causalist and non-causalist, to understand Aristotle as having held a theory of action. The possibility of contradictory interpretations on the matter may be further grounds for dismissing our present aim, but I will be arguing that, despite all of these initial doubts, *some* interpretation of Aristotle as a philosopher of action can indeed be grounded in the texts themselves. Thus, in what follows, I consider a variety of philosophical lenses through which we might understand Aristotle as having held a view about action, despite the fact that he never explicitly discussed the concept of intention or intentional events.⁷² I begin with possible causalist claims interpreting Aristotle as the originator of standard causal views, followed by an alternative, non-causalist interpretation of Aristotle's philosophy of action. I will consider the causalist interpretation to be specious at best, although it is not without its merits and partial overlap with an Aristotelian view of agency. The general aim in canvassing a variety of views is to offer at least some hope that the causalist interpretation cannot be the correct one, much less our only option for, understanding Aristotle as some kind of action theorist.

2.1 Was Aristotle an action causalist?

Many contemporary philosophers of action subscribe to causalism, which I have presented⁷³ as the view that action is defined as the event of a physical movement (i)

⁷² Like his ancient predecessors, Aristotle does deal with the concept of *praxis*, which denotes the concept of action or conduct, and we may think that it comes closest to the concept of the intentional. However, Aristotle's remarks on the concept are more in light of its contrast from *poiesis*, which in our contemporary schema also counts as intentional, and thus it is arguable that the concept of *praxis* alone does not narrow down to the concept sought by theories of action (*Nic. Eth.* 1140b 1-5). His discussion distinguishing *praxis* from *poiesis* does bear significance for the idea of purposiveness as characteristic of action, especially where the agent's ends are achieved not merely whenever an external product is completed (as is the case for *poiesis*), but also in the doing of whatever the agent seeks to do (as for *praxis*).

⁷³ See Chapter 1 for an explication of this definition of causalism.

that is caused in the right way by another event or state, and (ii) whose cause is, specifically, something mental or psychological.⁷⁴ Causalists adopt a variety of combinations involving these two elements: some are not event theorists, and thus embrace (ii) but not (i), while others accept both.⁷⁵ Causalist philosophers, some of whom are also influential interpreters of Aristotle, include the earlier David Charles (1984), Jennifer Hornsby, Alfred Mele, and, earliest and most influentially among them, Donald Davidson. Some of these explicitly argue that Aristotle defended a causal theory of action, while others take the weaker position of suggesting that Aristotle serves as one inspiration for modern causalism, and thus can be named as an ally of causalist approaches.⁷⁶ Davidson, for instance, seems to regard his action causalism as standing substantially in consonance with an Aristotelian approach to agency.⁷⁷ Given that the aim of Davidson's enterprise is to establish a causal relation between reasons – defined as the mental states they comprise – and the physical movements they produce, we may find that his attempt to associate Aristotle with his particular causalism surprising, for there is no evidence that Aristotle specifically conceived of *reasons* in terms of mental states.⁷⁸ Nevertheless, Davidson writes: '[T]he

⁷⁴ As I mention in Chapter 1, causalism takes on a variety of forms, however: some causalists believe that the cause must be antecedent to the physical movement, while others believe it could be simultaneous. For the present chapter, I will assume that a general characterization of causalism suffices for the discussion.

⁷⁵ It should be noted that the acceptance of (ii) but not (i) is compatible with the idea of agent causation.

⁷⁶ That is to say, the claim that causalists have taken this view of Aristotle is a very general one: it may mean as little as citing Aristotle as an ally in defending their causalist views. Aguilar and Buckareff, for example, offer a summary of 'Aristotle's account of the etiology of action inherited by later causal theories of action', and claim that 'to the extent that Aristotle had a theory of action, his theory is clearly a progenitor of the CTA [Causal Theory of Action]' (2010: 3-4).

⁷⁷ This is not to say, of course, that Davidson saw all of his views on action as corresponding to Aristotle's: he disagreed radically with Aristotle over the latter's conception of *akrasia*, for instance.

⁷⁸ As I will mention further on, Aristotle does talk about thought and desire as being involved in movement, but the connection between thoughts and desires qua mental

best argument for a scheme like Aristotle's is that it alone promises to give an account of the 'mysterious connection' between reasons and actions' (1980: 11). On Davidson's interpretation, the causal relation promised by Aristotle's account specifically involves the concept of a reason which, as the cause of action, can be broken down into compound mental states, a belief and a desire (or, in lieu of 'desire', what Davidson calls a 'pro-attitude').

Causalism, as Davidson adheres to it, appears to have the following main components or conditions: (c1) a causal relation between an agent's reasons or intentions⁷⁹ and his intentional movements, (c2) analysis of reasons⁸⁰ in terms of compound psychological states, viz., beliefs and desires, and (c3) action identified (or defined) as the physical movements produced by the psychological states, which constitute the agent's reasons for acting, and without which such movements would not count as action. There may be variations on the view among causalists, but let us presume that overall I have given causalism an adequate characterization in these terms. What then is the evidence for thinking that Aristotle inspired or indeed might have subscribed to such a view, either in its totality or at least in part?

To answer this question, it is instructive to revisit the oft-cited query proposed by Wittgenstein: what is left over when I subtract the rising of my arm from my having raised it? From a causalist perspective, the question seems to suggest the necessary conditions for action: the view that action requires the causal combination of certain elements. On the causalist view, for a movement (conceptually understood as 'mere

states as constituting *reasons for acting* are far more tenuous. The prime reason for doubting that the latter is an Aristotelian thesis is that Aristotle simply had no word that translates as 'mental'.

⁷⁹ Davidson, for instance, makes equivalent an agent's primary reason and the intention with which something was done: 'To know a primary reason why someone acted as he did is to know an intention with which the action was done' (1980: 7).

⁸⁰ In light of the definition of reasons offered here, it should be clear that the discussion henceforth focuses on motivating rather than normative reasons.

movement') to qualify as action, it must be caused by the appropriate psychological states, namely a desire and a relevant belief. Therefore, something further must be added to the concept of mere movement (sometimes described as 'bare movement'), for it to count as an action, such as my raising my arm. Based on this reading, the causalist's reply to Wittgenstein's question would be: what is left over from the conceptual subtraction is none other than the requisite mental states which stand in the proper causal relation to one's movements. If this reading is correct, however, then it is puzzling that one would say that modern causalists adopt the view that the rising of my arm is identical to my having raised it. For the causalist asserts that there can be risings of one's arm that do not amount to one's having raised it: such is the case wherever the necessary and relevant 'causal etiology' involving mental states or events is absent. So it cannot be that a mere rising of the arm is identical to the agent's having raised it, for the necessary and relevant mental states or events must be added to the mere movements to produce actions, such as someone's raising her arm.

It is perhaps even more puzzling that some commentators of Aristotle would claim that Aristotle holds a view, in common with the causal theorist, that the rising of one's arm is identical to one's having raised it. For although it may be said that Aristotle, like the causal theorist, identified the existence of certain causes of action, would Aristotle not have recognized that mere movements are precisely *unlike* actions, at the very least with regard to their respectively distinct causal character?⁸¹ However, Coope, for instance, adopts the very thesis that both Aristotle and the causalist are committed to identifying movements with actions (2007). She explains:

⁸¹ As I argue later, the source of these difficulties can be traced to the fact that many philosophers, such as Coope, omit to clarify from the outset that there may be important differences between types of 'movement'. The way that I wish to resolve the difficulty is by appeal to a kind of action disjunctivism. As it is, the notion of 'movement' is by itself too murky to be employed without more careful clarification, which would help to prevent certain difficulties arising as they do here.

The standard causal account, like Aristotle's, identifies my action of moving my body with my body's movements. So according to both accounts, when I raise my arm, my arm's going up is the same as the event of my raising my arm. Moreover, Aristotle, like the proponents of the standard account, holds that human and animal actions must be caused, in some sense, by a combination of the agent's desire and the agent's being in some cognitive relation to the object of this desire (2007: 110).

On Coope's interpretation, Aristotle agrees with the causalist as regards (c1)* that there is some causal connection between movement and psychological states, and (c2)* that the states in question are a combination of the cognitive and conative elements in sensing and thinking creatures.⁸² Insofar as such states can be implicated in explanations of the agent's movements, it is even possible to construe Aristotle as being in agreement with regard to these states constituting reasons for acting. Where Aristotle and the causalist apparently diverge lies in the third tenet of causalism: (c3) on the causalist's account, physical movements are identified as action in virtue of their being produced by beliefs and desires. The primary support for this claim is *De An.* III.9-11, where Aristotle cites both the causal contribution of desire and *nous* (which includes imagination) in the production of agential self-movement.⁸³ But Aristotle's account of action is not that movement is necessarily made into action by the psychological states of the agent, but merely that action *can* be explained with reference to such psychological facts; that action is thus explicable does not entail that the occurrence of such psychological states makes it the case that a movement is an

⁸² The asterisks signify modifications to the causalist premises (c1) and (c2), without which I believe Coope's interpretation could not be sustained.

⁸³ He says at *De An.* III.9-11 that 'these two, appetite and practical thought, seem reasonably considered as the producers of movement; for the object of appetite produces movement, and therefore thought produces movement, because the object of appetite is its beginning' (433a 19-20).

action.⁸⁴ Despite their differences on this last point, however, the Aristotelian and causalist views of action appear - at least on Coope's interpretation - to have a good deal in common.

In support of the view that these commonalities are legitimate, causalists unsurprisingly adopt a partisan reading of Aristotle's general statement that 'every action must be due to a cause' (*Rhet.* I).⁸⁵ On Davidson's view, there can be no way to understand this thesis other than in the light of *NE* VI.2: 'An origin of action – its efficient, not its final cause – is choice, and that of choice is desire and reasoning with a view to an end' (1139a 31-32).⁸⁶ Interpretation of the latter statement, in itself, is obviously not constrained by any one (much less all) of the propositions defining causalism, as characterized by (c1)–(c3). But for Davidsonian-type causalists, claims concerning action's 'origin' necessarily point to its principal causes (even supposing that final causes were recognized in the causalist's metaphysical scheme), where the cause of action is a choice made by the agent, understood in terms of deliberation and wanting. A request for an explanation on behalf of the agent as to why this or that was done by him will naturally refer to his object of desire and reasoning with a view to obtaining it, or perhaps just one of the two, and sometimes neither – for as Davidson remarks, 'it is generally otiose to mention both' (2001: 6). For the causalist interpretation to succeed, explanation of action in terms of beliefs and desires must

⁸⁴ Making the further case for Aristotle's alternative, of action as a change exercised by capacities, is the subject of succeeding sections.

⁸⁵ Notably, this statement of Aristotle's is consistent with an agent-causal view, i.e., one denying that the causal contribution towards an action is necessarily an event.

⁸⁶ Many translations render this passage erroneously: 'The origin of action – its efficient, not its final cause – is choice'. Aristotle's text merely says that choice (*prohairesis*) is an origin of action, among others that may not include the deliberation characteristic of *prohairesis*. At *NE* III, 112a 14 ff he writes: 'Not everything voluntary is chosen'. On this basis one can understand his view that non-rational animals and children can also be (voluntary) agents. See also *De An.* III.10 for Aristotle's more extensive argument that both appetite and mind are capable (and necessary) for causing movement.

make possible the critical linkage to *reasons*. The causalist places all emphasis on the claim that an origin and cause of action is agential choice, intention, or his reason for acting, which are not merely explicable in terms of psychological states such as wanting and deliberating, but indeed (in virtue of the causal relation established by the causalist reading of action's origins) a metaphysically requisite component for an event to qualify as an action. On this interpretation of *NE* VI.2, all three conditions of causalism indeed *appear* to be satisfied in this way.

2.2 Objections to the causalist interpretation

The foregoing causalist interpretation of *NE* VI.2, however, should be regarded as an over-simplification of Aristotle's thought. First, the causalist interpretation does not take sufficient account of cases in which the agent's reason for acting in fact specifically excludes the relevant desire. Aristotle observes in *De Anima* III.9 that 'it is not appetite which is responsible for movement; for the self-controlled, though they may crave and desire, do not do these things for which they have an appetite, but follow their reason' (433a 9 ff). By itself, this comment does not necessarily vitiate the causalist's commitment to a belief and desire pair as constituting the primary origin and cause of action, since causalists could argue that acting against a specific desire does not exclude acting from a secondary desire, i.e., the desire to follow one's reason rather than one's impulse.⁸⁷ Still, to regard belief and desire as constituting the primary cause of action is an error: for the causes of movement are many, according to Aristotle in *De An.* III.10, and although those cited by the causalist are among these, neither desire nor deliberation are the *first* cause, but rather 'the object of appetite (for this, though unmoved, causes movement by being thought of or imagined)' (433b 11 ff). Indeed, relative to the truly primary cause that is the agent's desired object, desire and belief play more of an intermediate causal role; Aristotle describes desire as 'that

⁸⁷ For Aristotle, appetite (epithumia) is but one of three types of desire (orexis).

which both moves and is moved', i.e., a moved mover. The agent's object of desire, in contrast, is an uncaused cause, and therefore primary.

The causal theory of action entirely ignores this critically important discussion in *De An.* III.9-10. Given the over-simplification of their interpretation, causal theorists may therefore reasonably face the criticism that their account of action fails to assimilate the Aristotelian notion of final cause. According to this objection, defining 'reason for acting' in terms of psychological states qua mental causes of action neglects the possibility that an agent may cite a reason for doing P as 'for the sake of' achieving Q, or 'in order to' do Q, etc.⁸⁸ In citing reasons that express purposiveness or final causality, agents themselves reveal a non-causalist structure of action wherein one's physical movements are not necessarily propelled or pushed forward by mental states. Action is instead forward-looking, drawn towards a future state of affairs by agents themselves.

Causalists who identify as Aristotelians are not without a response for this sort of objection. Reasons cited in favor of doing P for the sake of Q are expressive of antecedent (and/or sustaining) psychological causes whose content is forward-looking or in reference to a future state of affairs. Expressions of intention are, after all, frequently incomplete explanations where reasons for acting are concerned (Anscombe 2000: §3), and especially in cases where 'it is otiose' to cite the mental causes of action, as Davidson says. The apparent difficulty is dissolved once the causalist concept of wanting is assumed: the latter is both a causal factor for action and justifies those who cite reasons of the teleological sort. In support of this view as Aristotelian, causalist interpreters might appeal to *De Motu Animalium*, VII: 'All living things both move and are moved for the sake of something, so that this is the limit of all their movement that for the sake of which. Now we see that the animal is moved by intellect, imagination, purpose, wish, and appetite. And all these are reducible to

⁸⁸ See Ginet (1990).

thought and desire' (700b 15-18). Thus final causality does not appear to exclude psychological – both cognitive and conative – causation. This indeed seems to be Aristotle's view, especially given his remarks on the multiple causes of action in *De An.* III.9-10.⁸⁹

But it is a significant further step to establish the truth of (c3) from the aforementioned passage in *MA* VII. The third causalist condition defines action as identical with physical movements so long as these are produced by the relevant psychological states, without which such movements would not count as action. In order to make an explanation involving mental states, such as beliefs and conative attitudes, into the essential definition of action, one would be taking philosophical liberties not warranted by Aristotle's texts. At the same time, Aristotle does not deny that desire and intellect are somehow implicated in movement. Aristotle's view is that imagination, sensation, and thought are all faculties of discrimination, while appetite, impulse, and wish are all 'forms of desire, while purpose belongs both to intellect and to desire. Therefore the object of desire or of intellect first initiates movement – not every object of intellect, but only the end in the domain of conduct' (700b 20-25). It is clear from this that although desire and intellect in some sense 'initiate' movement, it is not necessarily as Davidsonian psychological states that they do so.

In light of the necessity condition involving psychological states in (c3), causalists are commonly accused of creating a problem known as 'the disappearing agent' (see §2.1 in Chapter 1). According to this objection, defining 'action' exclusively in terms of causation by mental states results in an irreversible shift from agent to event, in which the sense of action as something that *I do* is lost within the picture of action as an

⁸⁹ It should be noted that Aristotle also recognizes teleology without cognition and desire (e.g., in plants). This broadens the sense of purposiveness such that it ultimately seems trivial to cite it in agency. The apparent problem this produces might be dispatched, however, by placing commonsense limits on how far to apply the very same concept of purposiveness to human action as well as natural movement or accidental happenings.

event caused by another event (or set of events). For this reason, although Coope affirms that causalism has much in common with Aristotle, she ultimately rejects the interpretation of Aristotle as a proponent or originator of causalism. Coope's objection is that the causalist formula entails the problem of disappearing agents: it fails to establish true agency, in the sense that the causes of action, beliefs and desires, are mental states, rather than the agent herself. She writes: 'What the standard account fails to recognize (on this view) is that an action is *an agent's* bringing something about. This is not something that can be reduced to one event's causing another. The standard account assumes that the only sorts of things that can be causes are events and states. Because of this, it cannot make sense of a person's doing something' (2007: iii). Coope's alternative concept of action relies on a reading of Aristotelian change in terms of the exercise of an agent's capacities, where the problem of disappearing agency clearly does not arise.

As the ground for dismissing causalism, however, the threat of vanishing agents ultimately fails to impress. It is certainly doubtful that the objection would persuade causalists. Davidson himself anticipates this kind of challenge and retorts, 'Why on earth should a cause turn an action into a mere happening and a person into a helpless victim? Is it because we tend to assume, at least in the arena of action, that a cause demands a causer, agency an agent? So we press the question; if my action is caused, what caused it? If I did, then there is the absurdity of infinite regress; if I did not, I am a victim. But of course the alternatives are not exhaustive' (1980: 19). Davidson's point is that the problem of disappearing agency is premised by false alternatives; to say that action requires mental causation is not yet to admit that free agency is abolished. Nor must the causalist view reject the intuition that one's action is indeed one's own. Davidson's causalist theory, in any case, seeks to provide a reductive analysis of agency, the whole point of which is to define agency in non-agential terms. His component approach to action thus seems immune from the critique involving the notion of disappearing agency.

2.3 Further difficulties

But is causalism really exempt from the worry of vicious regress? One species of causalism is volitionism,⁹⁰ which maintains that physical action is caused by internal volitions or willings, conceived as mental actions. Volitionism is certainly vulnerable to the prospect of worrisome (if not infinite) regress, for it appears that its method of explanation of action is to cite another action as cause. Although they differ over the concept of mental cause, Davidsonian-type causalism⁹¹ bears family resemblance to volitionism in two key respects. First, both theories accept a conceptual separation between physical action as a movement, and the causes of that movement that result in it being an action. Causalism accepts these causes as mental or psychological states or events, while volitionism speaks of willing or volition as a separate event that moves or pushes the body into movement. In this respect, both theories presume a kind of dualism between mental and physical aspects of action. Second, both causalism and volitionism implicitly require a kind of introspection which amounts to empiricism about action: explanation of action searches inside the agent for the sources or springs of action.⁹²

⁹⁰ Perhaps not *all* varieties of volitionism are to be grouped together as sub-species of causalism. It all depends on whether one views ‘volitions’ to be separate events from the actions they cause, or rather constitutive of these events. O’Shaughnessy holds the latter sort of view, but arguably would not describe himself as a causalist.

⁹¹ This is not to exclude other variants, such as neo-Humean causal theories, of which Michael Smith is a notable proponent.

⁹² One may be tempted to add the idea that both causalists and volitionists require a key element in pushing or impelling a physical movement forward: the willing or mental states (which for Davidson simultaneously constitute reasons) by which a physical movement is also an *action*, i.e., a ‘free and intentional’ event. McDowell, for instance, accuses Humean-influenced causal theories of ‘adherence to a quasi-hydraulic conception of how reason explanations account for action. The will is pictured as the source of forces that issue in the behavior such explanations explain. This idea seems...a radical misconception of the sort of explanation a reason explanation is’ (1981: 155). Smith dismisses this criticism as involving a ‘prolix’ description of neo-Humean theories (such as Smith’s), where such theories offer,

Some volitionists respond to the criticism of regress by asserting that volitions are uncaused causes.⁹³ This is precisely Davidson’s response as well: mental causes, on his view, are ‘agentless’ (or not caused as by agency), thus apparently evading the criticism of vicious regress (1980: 19). Whether or not this rebuttal succeeds in either case, it is clear that the threat of regress is vicious precisely because the mental events cited as causing physical movements, whether volitions or reasons, themselves exhibit movement of the kind under scrutiny (and which require further explanation). For similar reasons, Aristotle was critical of Platonic views concerning self-movement, which appear to adopt an ancient form of volitionism (*De An.* 406b 26-28).⁹⁴ These accounts occur in both the earlier and later works, for instance in Plato’s *Laws* (894e-95a) and *Phaedrus* (249c-249d), both of which describe the soul as a mover of the body, and the soul itself as being moved by its own movement: self-movers produce motion by undergoing ‘self-generated motion’ in the soul (*Laws* 895a). Thus, an agent that causes change (i.e., towards another state) must itself be undergoing change: to cause

simply, a *causal* explanation of mental events with regard to action. So Smith does not entirely dismiss the accusation for being inaccurate; he himself writes: ‘To be sure, one who holds that reason explanations are causal must conceive of some psychological states as possessed of *causal force*’ (1987: 44). However, to claim that such causal theories share precisely the same volitionist element in which something pushes forward something else (as volitions cause actions) is likely unfair to the causalist, given that his conception of the causes of action is in terms of *states* (i.e., something stative), rather than volitional willing (i.e., something non-stative).

⁹³ See Schlosser 2012, ‘Agency, Ownership, and the Standard Theory’, for a summary of this response.

⁹⁴ Given his critical survey of earlier arguments, and depending on how volitions are understood, Aristotle might be construed as an anti-volitionist: see Whiting, Jennifer’s ‘Locomotive Soul’, *Oxford Studies in Ancient Philosophy*, 2002. Aristotle’s comments on the issue may be discerned in the following passage in *De Anima*: ‘Some say that capacity to produce movement is first and foremost the characteristic of the soul. But because they believe that nothing can produce movement which does not itself move, they have supposed that the soul is one of the things which move’ (403b 29-30). But this very much depends on whether one assumes that a volition is a movement. Aristotle’s views may be compatible with the idea that a volition is a non-movement type of cause of movement (as the soul is cause of animal movement).

a change, an agent must simultaneously be changing (Coope 2007: 116). But this cannot be Aristotle's view, since Aristotelian self-movers are at least partially unmoved while simultaneously causing movement.⁹⁵

In light of these ancient debates, it is evident that volitionism – as well as some varieties of causalism, to the extent that these theories share in certain characteristics of volitionism – suffers from a classical fallacy: the thought that the origin of a mover's manifest motion is a more primary (or internal) form of motion. Volitionists characterize agential self-movement or change in terms of an internal willing, while causalists identify the internal motion in terms of desiderative attitudes ('pro-attitudes'). It is important to distinguish this latter view from the idea that psychological causation is not incompatible with free agency. For Aristotle himself seems to suggest that such causation is perfectly compatible with the sense of agency which critics invoking the 'disappearing agent problem' wish to preserve, when he says at *NE* VI.2: 'Hence choice is either desiderative reason or ratiocinative desire, and such an origin of action is a man.'⁹⁶ Causal theorists including Davidson will not disagree with the latter statement, while nonetheless rejecting the notion of agent as cause. Instead, they insist that the causal constitution of action requires further explication, specifically, in terms of psychological states. Since such belief-desire compounds also count as an agent's reasons, causalists argue that they are thus exempt from the threat of vicious regress. So Davidson writes: 'Some causes have no

⁹⁵ A more precise statement of Aristotle's view on this matter should take note of his remarks at 202a 2-202a 11, where he explains that a mover is also moved accidentally, because movement is essentially of the moved qua movable. This is to say: the mover is 'per se unmoved' (while simultaneously happens to be moved) insofar as its operation as mover is the actuality that completes the movable thing's potentiality for motion. The text also expresses the simpler point that a physical agent, in touching the patient that it moves, is in turn moved. The movement by reaction of the physical agent is, however, only accidental to its action as a mover (its movement is a consequence of that action on the moved patient).

⁹⁶ One should read this passage as being consonant with *EE* II.6, where Aristotle refers to man as a point of origin (*archê*).

agents. Among these agentless causes are the states and changes of states in persons which, because they are reasons as well as causes, constitute certain events free and intentional actions' (1980: 19). This remark may set the causalist at a distance from volitionism, since willings always have agents; whereas Davidson can say that a desire coupled with a belief simply moves the agent, without requiring further explanation about how such events originated. But if mental causes are 'agentless', then it may appear to some critics that they are events that simply occur and which result in physical movements that we call action. That, at least, is the worry of the disappearing agent expressed by some philosophers, although it is not as clear that the worry has a substantial foothold here: for Davidson's point seems to be merely that mental events are not normally caused by agents, not that they have no causal explanation for their occurrence whatsoever.

As earlier discussions concluded, the prospect of disappearing agents in general may not be as pernicious as the critique against the causal theory maintains. Whether or not Davidson is susceptible to this problem, however, it seems advisable that causalism might be rejected on other, stronger grounds.⁹⁷ As understood in the foregoing, the Davidsonian-type causalism offers no reason to believe either that (c2) mental states such as belief and desire are simply equivalent to an agent's reasons, or (c3) that it is *in virtue of* the fact that action can be understood in terms of wanting and reasoning, that therefore certain movements count as actions. Aristotle's view is that the concept of action *can* be (and is) explicated in terms of desire and thought:⁹⁸ but it is not, on his account, that such mental states *ipso facto* necessarily ground the concept of action, nor that they always constitute an agent's reasons for acting. Causalists who claim Aristotle as the originator of their views therefore stand accused of promoting, at best, a specious interpretation of Aristotle. At worst, proponents of

⁹⁷ These grounds are already reviewed more thoroughly in my first chapter.

⁹⁸ On my reading of this, I have specified the modified premises as (c1)* and (c2)* in the foregoing.

the standard causal view, with Davidson at the forefront, write as if causalism is simply a true depiction of human action, or that their view *prima facie* has deep historical roots grounded in the Aristotelian tradition. But there simply is no good evidence for either of these assumptions.

3.1 Action as Aristotelian change

Proponents of the causal theory of action sometimes claim that Aristotle represents the philosophical progenitor of their view: their claim is not entirely groundless. As the foregoing has shown, it is possible that Aristotle agrees with the causalist as regards (c1)*, that there is some causal connection between movement and psychological states, and agrees with (c2)*, which claims that the mental states in question are a combination of the cognitive and conative elements in sensing and motile organisms. Nevertheless, the standard causalist view fails to adequately establish a connection to Aristotle as being substantially derived from his views on action (in addition to being rejected for the several independent reasons summarized in Chapter 1).

A different approach to developing an Aristotelian theory of action lies in the concept of action as Aristotelian change, a proposal favored by Coope (2007). On Coope's view, the problem of the disappearing agent (as canvassed in Chapter 1, §2.1) poses a significant challenge for the causal theorist. The problem arises from the causalist's reductive concept of agency, which posits that

when I raise my arm, all that happens is that my beliefs and desires cause my arm to go up. But in that case, so the criticism goes, I don't really *do* anything: I am merely an arena in which some events (my beliefs and desires) cause others (the movements of my body). What the standard account fails to recognize (on this view) is that an action is *an agent's* bringing something about (2007: 110-111).

According to Coope's interpretation, Aristotle's non-reductive view of action avoids this difficulty, for 'he does not share the assumption that the only things that do any causal work are events and states' (2007: 111). Rather, Aristotle's view is that an event is an action if it is the exercise of some kind of causal power. Coope's approach is to retain the general intuition that action is 'a causing of something', while rejecting the causal theorist's stipulation that the causation must be a relation involving (a certain conception of) beliefs and desires, where the occurrence of such mental states constitutes a necessary and sufficient condition⁹⁹ by which an event counts as an action (2007: 110). Her argument begins with the following key premises:

P1: An agent's action of moving her body is identical to her body's movement¹⁰⁰
(e.g., her moving her arm is identical with her arm's rising).

P2: A change counts as an action if and only if it is the exercise of a certain kind of causal power (e.g., an agent's arm going up counts as an action, viz., her raising her arm, only because it is the exercise of one of her causal powers).

Coope's proposal may be summarized in the following. If these two premises really amount to a position sustained by Aristotle (or one that he might have accepted), then Aristotle's view avoids the problem of disappearing agency; simultaneously, from these two premises it is possible to develop a full-fledged theory of action attributable to Aristotle. For on his view, Coope argues, an action is a change that only an agent

⁹⁹ As discussed in the preceding chapter, the other condition concerns the causal relation between the movement(s) and the mental states: the body movements must be 'caused in the right way' in order to count as action.

¹⁰⁰ More will be said on this later, but it appears that, for Coope, such movements are by definition intransitive. I believe this limit on the notion of a bodily movement is a mistake. For this reason, her account is unable to make coherent Aristotle's examples involving transitive actions dealing with external objects, such as housebuilding: that the bricks and mortar are a part of the action and yet external to the agent's movements are later raised as a puzzle for the Aristotelian view.

brings about: contrary to the causalist's understanding, Aristotle's concept of action is not reducible to the causation of one event by another. Since an agent's raising her arm is not something that can be reduced to a set of mental or psychological events that cause her arm to rise, the concept of action as something belonging to an agent is thereby preserved in the claim that to act just is to move one's body. For this latter reason it is important, according to Coope's interpretation, to emphasize (in the formulation of P1) that the action of moving one's body is identified with the motion of the agent's body. Such movement is the exercise of a causal power of an agent, a kind of agential occurrence that Coope argues is resistant to causal reduction. The resulting picture of agency, if the latter aim of resisting causal reduction can be achieved, provides a substantial contrast from the causal theory of action. Granting Coope's premises, it is then possible to define action (referred throughout this chapter as 'action as change') in the following terms:

Action as change: The exercise of a certain kind of causal power of an agent, also understood as the partial fulfillment of that capacity (in Aristotelian terms, 'the incomplete actuality of the moveable').

Coope then raises the following question: 'How *can* my action of raising my arm be the exercise of a causal power, if it is itself a movement my arm undergoes?' (2007: 111). That is, Coope's two premises taken together appear to produce a conflict for the definition of action as change, since they entail the absurd consequence that the causing of some event is identical to the event caused. If the action of someone's raising their arm is the causing of a change, namely, their arm's rising, and if their raising an arm is identical to its rising, then it does look as if we are forced to say that what is caused is identical to the causing of it. Coope's solution is to revise the first stipulation leading to the puzzle, by suggesting that 'an action is the causing of a state, rather than the causing of a change' (2007: 112).¹⁰¹ The revision leads to the following

¹⁰¹ 'State' here is understood to denote the *end state* of the movement.

clarification. An action therefore does not *cause* a change or a movement: instead, it *is* the change wherein the movements of one's body are identified with the action itself. Put yet another way: action causes not the changes that occur to a certain state of affairs, but rather the end state itself of those affairs. Action is the causing of the end state – it causes the end state to be – rather than of the changes that result in the end state.

The background to Coope's clarified definition of action lies, unsurprisingly, in Aristotle's view of change as the partial or entire fulfillment (or actuality) of a capacity (or potentiality) for the end state, where the concepts of potentiality and actuality are not opposed but rather conceptual supplements to one another. The implication of this notion of change for agency is that an agent's arm rising is thus identifiable with her raising it, because these events are not two separable components of an action. An agent's arm going up is a fulfillment of her capacity to raise her arm, rather than an effect of her action which caused it. The fulfillment of the agent's capacity is for an end state of affairs – that her arm is raised – and not, crucially, the fulfillment of the agent's capacity for *becoming* some end state. For such 'becoming' is a change, which would render Aristotle's definition of change, if given in such terms, circular.

If Coope's solution to the apparent puzzle succeeds, then her definition of action as change is able to resist the absurd consequence that the causing of some event is identical to the event caused. Its definition of action, she suggests, is also potentially useful in better understanding Aristotle's argument concerning self-movement in *Physics* VIII.5. There he writes:

Movement is an incomplete actuality of the moveable. But the mover is already in actuality, e.g. it is the hot thing that produces heat, and in general, that which produces the form possesses it. So that the same thing in respect of the same thing will be at the same time both hot and

not hot. ... Therefore, when a thing moves itself, one part of it is the mover and another part is moved (257b 8-13).

For Coope, the puzzling aspect of Aristotle's argument lies in the apparent claim that 'what changes (or moves) something towards being F must itself be F (e.g. something that heats must be hot)' (2007: 114). Coope argues that the only way to make sense of this premise is to interpret Aristotle as specifically resisting the Platonic thought that what changes something towards being F must itself be undergoing change. The latter Platonic idea is an assumption of many contemporary volitionist approaches, as well as some varieties of the causal theory of action. The acceptance of the assumption entails that what is caused in the fulfillment of the self-mover's capacity is a change rather than the being of the state. As discussed in the preceding section, however, Aristotle is critical of this kind of view of self-movement (*De An.* 406b 26-28). So the foregoing counts as one reason to interpret *Physics* VIII.5 as requiring that the action of changing something towards F is a causing of a state (*being F*) rather than the causing of a change.

Coope remarks that this notion of 'causing' has no obvious equivalent in Greek (2007: 113). It seems plausible, nevertheless, to ascribe the concept of causing to Aristotle's idea of movement as an incomplete actuality of the moveable; in the case of action, the agent as mover causes (or fulfills) a potentiality for being in some end state, and the action itself is the incomplete actuality of what is moveable (the agent's body). But this definition requires us to characterize action as something 'incomplete'. Such terminology no doubt sounds alien to modern ears, and demands clarification. If it can be made intelligible, how does it ameliorate the puzzling claim, raised earlier, that what changes or moves something towards being F must itself be F ? In *Physics* III.1,

Aristotle defines change as the actuality of what is potentially¹⁰², so the answer will turn on the significance of the phrase, ‘what is potentially’, as it functions in that definition. Coope presents two possible interpretations of the phrase. On the first interpretation, change is the actuality of what is potentially changing, *qua* such (Heinaman 1994). On the second interpretation, change is the actuality of what is potentially in an end state, *being F*, *qua* such (Kosman 1969, Waterlow 1982, Coope forthcoming).¹⁰³ Only the latter is able to make sense of the idea that action is something incomplete, for by the second definition, action is a movement towards a state that has yet to be achieved, namely, *being F*. If this analysis is correct, she argues, then it is possible to resolve the puzzle that what is moving something towards being *F* must itself be *F*, or, as Aristotle puts it, ‘the same thing in respect of the same thing will be at the same time both hot and not hot’ (257b 8-13).¹⁰⁴ Coope offers the example

¹⁰² Coope’s translation is as follows: ‘Change seems to be a kind of actuality, but an incomplete one. This is because the potential, of which it is the actuality, is incomplete’ (201b 31-33).

¹⁰³ Coope identifies a third possible interpretation, on which change is ‘the *actualization* of what is potentially *F*, *qua* such’ (Kosman 1987). It’s not clear, however, how this interpretation really distinguishes itself from the first available interpretation, so I disregard it here.

¹⁰⁴ Coope potentially faces a problem of interpretation throughout this discussion, i.e., it is possible that her reading over-complicates Aristotle’s passage at *Physics* VIII.5. The apparent puzzle that she identifies from the passage only arises if one assumes that Aristotle is interested in analysing the concept of change *tout court*. However, Aristotle’s remarks have a very specific target, as they are prompted by a question concerning self-movement: ‘If there is something which causes itself to be moved, how does it do it and in what way?’ (257b 1ff). So his point that ‘the same thing in respect of the same thing will be at the same time both *F* and not *F*’ is meant to illustrate that the nature of motion is such that nothing can move (or change) itself without a mover. Something cannot be both hot and not hot all by itself, but rather it is the mover that is hot and the moved that is not hot: because the mover and the moved ‘share the same name’, both being in respect to the same thing (the motion of heating), Aristotle’s apparently contradictory statement is clarified as arguing for a distinct mover and a moved (or a changer and a changed) in all motion or change. This point is distinct from the question of how ‘is *F*’ and ‘is not-*F*’ both in some way

of an acorn and its change into an oak tree to illustrate Aristotle's apparently contradictory statement:

The acorn is potentially an oak tree. Before the acorn has begun to grow, this potential is not, in any sense, fulfilled. When the acorn has finally become an oak tree, its potential to be an oak tree is completely fulfilled, and it can no longer be described as 'changing into an oak'. The acorn is becoming an oak just when its potential to be an oak is *incompletely* fulfilled. A change, on Aristotle's view, is essentially a going towards some state. In defining change as the incomplete actuality of a potential to be in some end state, he is saying that changing to *F* is an incomplete way of fulfilling a potential to be *F* (2007: 119-120).

Action, on her view, is an Aristotelian change wherein the self-mover, at the moment of acting, is both potentially in the state of having his arm raised, and also actually raising his arm. So it is possible to say, as Aristotle suggests at *Physics* VIII.5 with regards to heat, that the agent simultaneously does not have his arm raised (since he is potentially in the state of having it raised), but that he is also raising his arm (since he is actualizing his potentiality for that state).

In sum, Coope's proposal is that we can understand Aristotle's view of active change as a fully-fledged theory of action, in which 'the power that is exercised in an action of *moving X* is a power to produce the *end* of *X*'s movement: a power to produce a state, rather than a movement' (2007: 113). Compared with the previous interpretation of Aristotle as a causal theorist, the concept of action as Aristotelian change indeed seems far more plausible. Coope's analysis is not without potential difficulties, however. To these I turn next, in the following section.

true of the thing moved, where the solution includes the qualification that the thing moved is, in actuality, not *F*, but is, potentially, *F*.

3.2 First difficulties: ‘change’ and natural phenomena

The first problem with Coope’s account of action as Aristotelian change is one that she herself recognizes. She considers that her account may be ‘criticized on the grounds that there are certain things that we would, intuitively, want to count as changes that are not...directed towards an end’ (2007: 135). The concept of action under consideration has been in terms of Aristotelian change, an incomplete actuality of a potential for an agent’s body (or part of her body) to be in some end state. This definition appears to entail that ‘[f]or something to count as a change, then, it must already exhibit a kind of directedness: it must be a *going towards* some end’ (2007: 135). But one can think of plenty of instances of natural phenomena that should, it seems, be unproblematically identified as ‘changes’, and yet which do not appear to be ‘going somewhere’, i.e., towards some final end state. Coope gives the examples of the motion of a stone when it is accidentally rolled out of the path by a walker, or the erratic arm movements of someone suffering from alien limb (or ‘anarchic hand’) syndrome. If there is a basic sense in which ‘change’ is understood as the motion of some physical body, then why should such occurrences not count as changes?

Coope’s response to this problem is essentially that an Aristotelian action theorist must ‘bite the bullet’ and reject the rolling of a stone from its path, or the involuntary movement of an arm, as real changes. Her solution re-emphasizes the interpretation that a certain concept of intrinsic directedness must be present for a physical movement to qualify as a true Aristotelian change. She writes:

A movement of this kind [i.e., the erratic arm movement of a patient with alien limb syndrome] is not the fulfilment of a potential anything has to be in some particular end state. If it is right that such accidental movements do not qualify as Aristotelian changes, then this has an interesting consequence for his views about action. On Aristotle’s view, there is not one type of event, *my arm’s going up*, which could occur

because I suffer from anarchic hand syndrome. The event which occurs when I raise my arm is of quite a different kind from the event that would occur if my arm moved about because I suffered from this condition. The one event is a change and hence a progression towards some end; the other is not. According to Aristotle, if my arm's rising is a change, it must have *some* agent (though the agent could be someone other than me, and it could, even, be something inanimate). On the other hand, if it is *not* a change, then it is not even a candidate for being identical with an action of the sort I have discussed in this paper: an action of changing something (2007: 135-136).

The above passage raises several interesting issues. The first is the idea that the question, 'What makes my arm's going up a case of my raising my arm?', does not even arise from an Aristotelian account of actions as end-directed changes.¹⁰⁵ Nor does Wittgenstein's question from *PI* §621 ('What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?'), from which the problem of action arguably derives, arise from Aristotle's schema. The reason these related questions do not arise for Aristotle is that his concept of action as a change implies that an action is an event 'that *simply in virtue of being a change* has a kind of directedness towards an end: it is the incomplete fulfilment of a potential of my arm to be up. Thus, in asking what makes this event an action, one is not asking what gives it this end-directedness (since this end-directedness is presupposed by its being a change at all)' (2007: 136).

The most obvious problem with this account is that most philosophers will not be willing to concede the claim that natural phenomena such as undirected or random rolling stones, and involuntary or non-voluntary behavioral events, do not count as changes. Most theorists will view as absurd a definition of change that excludes such

¹⁰⁵ Even if Aristotle were asked this question, remarks Coope, 'he could not understand this question in the way that it is generally understood in modern philosophy of action' (2007: 136).

occurrences. The costs of such a thesis simply seem too high, even if the thesis has the advantage of entailing that the problem of action (and Wittgenstein's question at *PI* §621) is rendered irrelevant by that definition of change. If the concept of Aristotelian change requires a kind of directedness towards an end, and since involuntary and non-voluntary behavioral events are said to lack such end-directedness, then involuntary and non-voluntary behavioral events are not changes, although they are movements of the body. Again, it seems unintelligible to identify an event as a movement but not a change.

3.3 Aristotelian change as a disjunctivist approach

The above passage from Coope presents, however, an interesting idea: that in distinguishing between an agent's action of raising her arm and the event of her arm involuntarily rising, a disjunctivist conceptual structure may be discerned in Aristotle's approach to action (defined as change). It is clear from Coope's picture that raising an arm voluntarily and allowing it to rise involuntarily are two very different events; a disjunctivist might identify them as being of 'fundamental different kinds'. It is also clear that 'there is not one type of event, *my arm's going up*', which is held in common by both voluntary and involuntary cases alike. That is, there exists no common factor, i.e., a mere bodily movement, that is shared by both the 'good' and the 'bad' cases. On the contrary, Coope's picture suggests that there are two fundamentally distinct senses of an 'arm-rising'. One could identify 'arm-rising_v' (a voluntary movement) as an incomplete fulfillment of an agent's potential to raise her arm: by Aristotle's definition, it is an event exhibiting end-directedness. The other kind, 'arm-rising_i' (an involuntary or non-voluntary movement), denotes a mere bodily movement, one lacking in end-directedness, for it does not fulfill any capacity or potential of an agent to move her body 'towards some end'.

As the discussion of disjunctivism in Chapter 1 suggested¹⁰⁶, the notion of a mere movement such as arm-rising_v is a critical assumption of the causal theory of action; it plays the role of a common factor shared by both the voluntary and the non-voluntary (or involuntary) kinds of movements. Assuming its presence in both cases allows the causal theorist to argue that what makes a movement an action is the addition of mental states or events: a bodily movement *is* an action if its ‘causal etiology’ reveals such an addendum.

It may be thought that Coope’s proposal agrees with the causal theorist on the latter point, i.e., that a bodily movement is identical to an action provided that certain mental states or events are present. For one of the premises of Coope’s account of action as change, as discussed earlier, is the claim that ‘an agent’s action of moving her body is identical to her body’s movement’ (P1). But if action is identical to a bodily movement, where ‘bodily movements’ may be understood as being either voluntary or non-voluntary, then one cannot deny the presence of a common factor shared by both voluntary and non-voluntary (or involuntary) behavioral events. The resulting picture is incompatible with the disjunctivist suggestion in Coope’s remark that ‘[t]he event which occurs when I raise my arm is of quite a different kind from the event that would occur if my arm moved about’ involuntarily. The discussion of disjunctivism in the preceding chapter showed that a disjunctivist approach to action must reject the notion of a common physical factor: one cannot have both.

The latter conflict turns on an ambiguity in the way ‘bodily movement’ is understood. It is not specified in P1 whether the body movement refers to a behavioral event of the kind expressed in arm-rising_v or arm-rising_i. If the disjunctivist suggestion in Coope’s interpretation of Aristotle’s concept of action as change is to be made consistent with what she claims in P1, then the kind of bodily movement cited in the latter premise

¹⁰⁶ I review a version of a disjunctivist approach to action in my first chapter, with the aim of defending it as a substantial alternative to action causalism, or the standard theory of action.

must be identified with the voluntary raising of one's arm, arm-rising_v. The type of mere bodily movement denoted by arm-rising_i cannot be involved in the claim that an agent's action is identical to her body movement (Pr), for the kind of movement denoted by arm-rising_i is fundamentally different from the kind of movement involved in Aristotle's concept of agency. For on Coope's interpretation of the latter, a movement is an action if it is an Aristotelian change, and many of the random natural happenings in our immediate environment – the rolling of a stone from the path, for instance – do not count at all as Aristotelian changes.

A problem with the above attempt to clarify the ambiguity involved in defining 'bodily movement' is that Coope does not adequately stress the contrast between her interpretation of Aristotle's view and the causal theory of action. It is clear from the quoted passage above that there are two distinct types of arm-risings, and only one of them (the voluntary kind) is implicated in Aristotle's explanation of action as change. But Coope claims that Aristotle's view retains 'much in common with the standard causal account' (2007: 110):

The standard causal account, like Aristotle's, identifies my action of moving my body with my body's movement. So according to both accounts, when I raise my arm, my arm's going up is same as the event of my raising my arm (2007: 110).

This claim may be doubted for at least two reasons. First, from the causal theorist's perspective, the causal etiology of a bodily movement is the determinant of whether the movement is identified as action or not. The Davidsonian causalist requires the presence of certain antecedent mental states or events, the latter which Aristotle's account recognizes as sometimes playing an important role in agency, but which are denied in his account as necessary conditions for a movement to count as action.¹⁰⁷

¹⁰⁷ Coope characterizes the disagreement between Aristotle and the causalist approach in the following remarks: '[A]ccording to the standard causal account, what makes the

Thus, if one wishes to say that the causal theory of action and Aristotle's account both identify action with a body movement, it must be clarified that these understandings of 'identity' are very different from one another. The causal theory claims that body movement is identified with action on the condition that certain mental states or events are antecedently present.¹⁰⁸ Aristotle's account, as Coope interprets it, holds that a body movement is identified with action in the event that the movement is the exercise of an agential capacity.

The second reason for doubting the earlier claim, that the causal theory and Aristotle's account are in agreement with respect to identifying the action with a bodily movement, can be drawn from Coope's own argument. At the conclusion of her essay, she states that Aristotle's conception 'of the event of *my arm's going up* is different from that presupposed in...modern accounts. On Aristotle's view, a change is the incomplete fulfilment of a thing's potential to be in some end state. For something to count as a change, then, it must already exhibit a kind of directedness: it

movement the agent undergoes an action of the agent's is the fact that it is caused in some appropriate way by the agent's beliefs and desires: my arm's rising counts as my raising my arm just because it is caused by my beliefs and desires. In contrast, though Aristotle thinks that human actions are explained by desires and beliefs, it is not *because they are explicable in this way* that they count as actions. On Aristotle's view, what makes a particular change *an action* is the fact that it is the exercise of a certain kind of causal power: it is a *causing* of something. My arm's going up counts as an action of mine (that is, it counts as my raising my arm) only because it is the exercise of one of my causal powers' (2007: 110).

¹⁰⁸ One reason for which Wittgenstein may have raised his infamous question about subtracting bodily movements from actions (2001: 136), was to suggest that contemporary thinking about action entailed a gap between action and movement, wherein a mental aspect is necessarily added to movement in order to produce action. The gap is illustrated not only by many causal theories but also some approaches that oppose causalism, such as Hornsby's characterization of action, on which the action of moving my body and my body's motion are two separate events, where the first event causes the second (2004). However the gap is formulated, the standard causal account of action sustains the idea that there can be arm risings without my raising my arm, where 'my arm rising' is understood as a merely bodily event barren of psychological or mental elements.

must be a *going towards* some end' (2007: 135). There seems to be a problem with reconciling this latter statement with the claim that Aristotle's account and the causal theory both share the view that actions are identical to body movements arises. For Coope herself, as has already been mentioned, states that the conception of the event 'my arm's going up' (where that event is the exercise of an agential capacity) is different for Aristotle than for the causal theory of action. The difference, which is not an insignificant one sustained between their concepts of *agency*, makes it impossible to claim that Aristotle and the causal theorist share the same view on any statement involving the concept of action (and its identity with the bodily movements of an agent).

Coope is therefore right to argue that the causal theorist adopts certain modern presumptions concerning the concept of 'physical movement', presumptions that Aristotle would not have accepted. I mentioned earlier that her argument for this point suggests differentiation between two fundamentally distinct types of physical movements, i.e., arm-rising, and arm-rising₁. One of these, arm-rising₁, is part of the conceptual structure of Aristotle's understanding of action; the other type plays no part in the concept of action at all. Coope's analysis of this disjunctivist picture is that one event, being the exercise of an agential capacity, is a change, and the other is not. Aristotle's view is considered problematic, as already suggested, merely on account of his definition of action as change. A further problem is introduced by the possibility that his definition of action provides the inspiration for a disjunctivist approach to action. The problem is that the disjunctivist approach claims a fundamental difference between types of events, but the conceptual basis for claiming the difference may be thought as tenuous at best (and unintelligible at worst). For according to Coope's interpretation, Aristotle understands 'one's arm going up', when it is an exercise of a capacity of the agent to move in such a way that her arm goes up, as the actualization of a potency, i.e., an event that intrinsically possesses end-directedness. The latter explanation of action had the benefit of justifying the claim that Aristotle could not

possibly have identified action with physical movement so as to agree with the causal theory of action, for the two accounts differ significantly over how the concept of physical movement ought to be understood. But this very explanation of action, i.e., that action is an event that intrinsically possesses end-directedness, faces a number of challenges.

3.4 Difficulties with ‘end-directedness’

The first difficulty is related to the problem already raised for Aristotle’s concept of action as change. The objection raised for the latter concept was that it absurdly excluded phenomena, such as the random rolling of a stone from a pathway, from the class of events that Aristotle’s definition recognized as changes. The concept of action as change also appeared to have the absurd consequence of entailing that some bodily movements are not changes. If Aristotle’s concept of change is so problematic, it may be objected, then it hardly serves as a promising basis for developing and defending a disjunctivist approach to action. The worry from the disjunctivist perspective is that the explanation of action in terms of Aristotelian change results in the following inept picture. If an event qualifies as an Aristotelian change, then it is an action (an intentional movement). If it is not a change, it is a mere movement, an involuntary or non-voluntary event. But even if this particular expression of disjunctivism counts as a sound interpretation of Aristotle, it surely poses a problem for the intelligibility of one’s theory of action.¹⁰⁹ My arm’s rising, although outside of my awareness and

¹⁰⁹ For Aristotle thinks that ‘all things existing by nature appear to have in themselves a principle of motion and of standstill, whether with respect to place or increase or decrease or alteration’ (*Physics* II.1). The things included in this definition are not only animals but also plants and simple bodies, such as earth, fire, air, and water. Coope’s substantiation for her interpretation is largely based on *Physics* III.3. The examples there that Aristotle offers on change, and with which Coope deals primarily, are housebuilding and teaching. Such actions take time, and often deliberation. But she simply equates these activities with the raising of one’s arm: both are subject to the same analysis. In addition, as Coope herself admits, the examples that Aristotle uses in *Physics* III.3 are also of inanimate things, such as fire.

intention that it rises, nevertheless is also a kind of change: doubtless it is a bodily motion.¹¹⁰ Therefore the prospect of construing such a motion as disjunctive from action, on the basis of the claim that one is a change and the other is not, is doubly implausible.

Another problem with the account of action as end-directed change is that, as Coope acknowledges, the event of my arm's rising counts as a change so long as it has 'some agent (though the agent could be someone other than me, and it could, even, be something inanimate)' (2007: 136). That is, my arm's rising might be brought about by another agent's external force applied to my body, and yet the resulting physical motion counts as an end-directed change – for even on this occasion, it is the fulfillment of my arm's potential to be in some end state, i.e., in a position of being raised. So in both the voluntary and involuntary cases here, there is an end-directed movement. Either it should be granted that both are actions¹¹¹, or some further strategy of differentiating them is needed. The question that potentially arises for Aristotle's account given this consideration is whether it is coherent to claim that there is such end-directed movement that exists in common between the voluntary case of my raising my arm and the involuntary hypothetical where external force is applied to make it rise. Although there may be possible solutions to this difficulty, Coope does not address them.¹¹²

¹¹⁰ And 'every motion is a kind of change', with the sub-types being three: that from a subject to a subject, from a subject to a nonsubject, and from a nonsubject to a subject (*Physics* V.1).

¹¹¹ In response to this sort of hypothetical, Davidson argues for the conclusion that they are both actions (2004).

¹¹² The most probable answer to this question, that Aristotle might have given, is no: that we cannot speak of a shared kind of movement (i.e., possessing end-directedness) between the voluntary case and the involuntary situation in which my arm is forced. One could argue that, for Aristotle, the raising of one's arm is a distinct type of movement that is an exercise of the body's instrumentality (or actualization of the physical potentiality) for the agent's specific ends. This argument could draw from the

Coope's explanation of the difference between action and mere movement raises further difficulties. According to Coope's explanation, the event of one's involuntary arm movement is not a change in the same sense that authentic action is a change: the latter is an exercise of a capacity of the agent.¹¹³ The critical question, known as the problem of action, is: what makes it the case that an involuntary arm movement is not the exercise of a capacity of an agent, whereas an agent's raising her arm is such an exercise of a capacity to do just that? A response claiming that the latter is an exercise of a capacity *to act* (or to achieve an end state *F* by acting) would be unsatisfying in its circularity. Coope's reply is that arm-rising is the kind of event that intrinsically contains end-directedness, or purposiveness. Coope does not offer much explication on the latter concept, besides stating that it is 'a kind of directedness...[that] must be a *going towards* some end' (2007: 135). Such teleological explanation will be familiar, even seemingly commonsensical; and yet, for many contemporary philosophers, it will also be unsatisfying.¹¹⁴ For until a precise meaning of an event or phenomenon 'going

definition of soul given in *De An.* II, that the soul is the first actuality of a natural *organic* body, where 'organic' is derived from *organon*, or tool (instrument). The living body is therefore the natural *organon* by which the soul or animal itself achieves its ends, whether voluntary or not. But by this definition we are able to distinguish between cases in which I raise my arm as a matter of 'using' my body to fulfill my end (of having my arm thus raised), and the artificial involuntary hypothetical in which the forcible raising up of my arm is, in contrast, merely an exercise of my arm's capacity to be dragged up, as any material body might be.

¹¹³ This also raises the more basic problem that the notion of 'change' may insufficiently characterize action, even for Aristotle, insofar as certain unintentional movements (e.g., shrinking away upon contact with heat from a fire) are built-in responses to stimuli which do seem to exhibit tendencies (in the latter case, namely, to preserve oneself from harm). Or consider the lower body movements that kick in when one stumbles over something in the path, the movements that occur to compensate from loss of balance. Such movements are not intentional, but nor should we say that they are *not* tendencies of an agent.

¹¹⁴ Employment of the concept of purposiveness or end-directedness suggesting a teleological structure of physical movement offers another reason for contemporary philosophers to reject Aristotle's view of action as change: explanation of action in terms of teleology is certainly rejected by the causal theory of action, if not most

towards some end' is spelled out, there will be potential confusion about what may be allowed as an application of that definition, and what ought to be excluded from it. Someone's raising her arm to reach for an object on the shelf clearly has an end in mind, or in sight, and her movement exhibits that 'directedness'. At the same time, her lungs expand as she breathes normally: a bodily motion, but not an action. On Aristotle's account, the expansion of the lungs is obviously not an action, but not because it is not 'going towards some end': for the expansion of the lungs in breathing fulfills a natural purpose (to oxygenate the blood). That is, the expansion of the lungs is a bodily movement with a very specific end: it is a movement that oxygenates an organism's blood; yet it must be distinguished from other end-directed movements, such as reaching for something on a shelf, that are end-directed movements that count as actions. What makes it the case that the expansion of the lungs in order to oxygenate the blood is *not* a 'going towards some end' type of movement that counts as action, whereas raising an arm to reach for an object *is*? The fact that one is an action and the other is not, based on the explanation that one is the exercise of an *agential* capacity while the other is a mere movement of a bodily process, is utterly unsatisfying; for if the latter is the only reply to the present ambiguity then the discussion is stalled because of circularity.

Perhaps these objections are unfair to Coope; perhaps she has in mind an explanation of action in non-agential terms that also goes deeper than merely positing that it is an event 'going towards some end'. But from what I can discern from the present essay, her disjunctivist interpretation of Aristotle appears simply to assert that there is a fundamental difference between action and mere movement, one kind of event that 'goes towards some end' and another kind that does not. If so, then Coope's approach is arguably subject to the same basic criticism of disjunctivism discussed in Chapter 1, i.e., that the disjunctivist approach fails to substantiate its main argument beyond a

contemporary philosophers tout court. This particular feature of Coope's account places Aristotle at even greater distance from the standard account of action.

facile denial of the causal theory of perception (and action). Disjunctivists claim a fundamental difference between types of experience: in the perceptual case, the difference is between veridical experience and illusory or hallucinatory experience, and in the case of action, the difference is between arm-rising_v and arm-rising_i, but the objection was that disjunctivism ‘makes no move *beyond* this at all’ (Dancy 1995: 435). In the preceding chapter, it was suggested that this objection to disjunctivism was not an insurmountable one.¹¹⁵ In any case, it seems probable that the view under consideration, of action as Aristotelian change, must answer to the same challenge raised by such critics of disjunctivism. Both arm-rising_v and arm-rising_i are bodily motions, and in both the veridical and illusory experience there is, at bottom, a ‘seeing an appearance’.¹¹⁶ The appeal to arm-rising_v as the causing of a motion that is purposive or end-directed is arguably analogous to claiming that certain perceptions are veridical because they are caused in the ‘appropriate way’. But as an attempt to explain the fundamental difference between experiential kinds, the disjunctivist approach thus far does not effectively extinguish the possibility that the concept of action might be less problematically explicated by the causal theory.

For a proponent of the causal theory may plausibly claim that what distinguishes the experiences ‘fundamentally’ is merely a difference in the causal relations working in the background of a perceiver’s veridical and non-veridical experiences, as well as an agent’s arm-rising_v and arm-rising_i. That is, it may seem as if the disjunctivist has distanced herself from a causal theory by baldly denying that veridical experience (or

¹¹⁵ As suggested in Chapter 1, a possible way out of Dancy’s objection lies in the disjunctivist’s appeal to the suffusion of inner and outer realms (Haddock and Macpherson 2011). What precisely that means, however, is yet unclear: the aim of the latter part of the present chapter is to clarify it using Aristotle’s concept of ‘enmattered formulae’, or hylomorphism in action.

¹¹⁶ The phrase ‘seeing an appearance’ is Anscombe’s (2000: 49). It seems to express very directly the notion that perception is necessarily mediated, although the latter need not be the case: it *is* only on the basis that one grants that appearances can be conceptually and ontologically separated from the reality they describe.

an arm-rising,) shares an essential constituent of an appearance (or an arm-rising;) in common with the illusory experience (or an uncontrolled arm movement). But the Aristotelian appeal to action as purposive or end-directed motion (versus non-purposive, directionless motion), does not in fact rule out the possibility that purposive motion might be partly constituted or produced from the very capacity by which one moves *simpliciter*, whether purposively or non-purposively.¹⁷ The present appeal to action disjunctivism, using the Aristotelian concept of action as change, thus fails to sustain itself as, most basically, a rejection of the causalist view.

These objections do not necessarily amount to the claim that the disjunctivist approach to action (as well as the original theory of disjunctivism with regards to perceptual experience), is without possible vindication. Under the current Aristotelian interpretation of action as change, however, the grounds for sustaining the latter's basic argument for the fundamental differences between experiential kinds appear to be inadequate in overcoming the criticism that disjunctivism fails to substantiate its primary argument. For the analysis offered by Coope presents another, more basic problem. Coope's argument is premised on her criticism of the standard account of action as promoting a 'disappearing agent', whose existence is reduced to the causal exchange between events or states, i.e., physical movements produced by mental states. Beyond providing a potentially plausible account of action disjunctivism, Coope's account of an Aristotelian action theory holds what she would view as a more important aim: it purports to preserve the concepts of agent and agency from reduction to the causal theorist's favored *relata*. It is not clear, however, that the Aristotelian account proposed achieves its end in this regard. The causal theory offers a reductive approach to explaining action and agency: specifically, it claims to explain these concepts in non-agential terms without undermining the intuition that action is

¹⁷ I do not mean to say that Aristotle himself would have explicitly recognized such a capacity: but under the current disjunctivist appeal such capacities could not be satisfactorily ruled out.

something done by an agent, rather than events and states. Particularly given that the Aristotelian view of action as change seems limited in *its* explication of action, what is to prevent a reductive explanation such as that offered by the causalist from being applied to Coope's account? That is, her account does not clearly show why Aristotle's account of the agent is not reducible to a concatenation of states or events. Perhaps there is a way out of this problem, specifically by a development of Coope's central notion of an event 'going towards some end'; but in its absence, causalists will argue that the account of action as Aristotelian change is better off, and more informative, under a reductive analysis.

To summarize, the account of action as Aristotelian change has certain limitations that render it inadequate for achieving the aims of this chapter, among which was cited the attempt to develop a characterization of Aristotle plausibly considered as engaging contemporary action theorists. The Aristotelian concept of action as change, i.e., as the actualization of a potentiality, is one that assumes an obscure notion of end-directedness in bodily motion from the outset. Subsequently it does not serve well as a conceptual paradigm for attempts to understand intentional movements, i.e. action.

4.1 Action as psychophysical process

That most philosophers would likely reject the account of action as change, as considered in the preceding, does not entail that it is impossible to discern a plausible disjunctivist approach to action from Aristotle's thought on agency. For there is at least one other possible interpretation that holds potential for being developed into a theory of action and is, in particular, one that may vindicate the thus far tenuous idea of disjunctivism apropos action. Having rejected the argument that Aristotle's concept of change by itself can provide the solution to the problem of action, i.e., 'What distinguishes action from mere movement (e.g., my raising my arm versus my arm

going up)?”, I turn to consider the interpretation, offered most recently by Charles (2008, 2011),¹¹⁸ of Aristotle’s concept of action as a psychophysical process.

Charles’ main aim is to develop Aristotle’s statement, in *De Anima* 3.10, that ‘desire moves the animal’ (433b 13 ff). Charles’ principal claim is that desire,¹¹⁹ along with other affections or emotions such as fear and anger, is ‘treated by Aristotle as inseparable in definition as well as in existence from states of the body’ (2011: 77). In the case of anger, there is a type of desire specifically for revenge, understood in accordance with the principal thesis: ‘[T]he type of desire for revenge which defines anger cannot be defined without reference to certain physiological processes’, which according to Aristotle, is constituted by a ‘boiling of the blood’ (2011: 77). Charles explains: ‘The [physiological processes] play a role in determining the type of desire that is present in anger. Desire of this type is an inextricably and essentially psychophysical process. It is a boiling-of-the-blood-type of desire for revenge’ (2011: 77). He also adds a further clause to his principal claim: ‘The type of desire for revenge which the angry experience resists decomposition into distinct physical and psychological components’ (2011: 77).

Charles’s interpretation of Aristotle, with regard to the principal claim that desire is ‘an inextricably and essentially psychophysical process’, is based on several key

¹¹⁸ Since Charles’s 2011 book chapter is, for the most part, a revision of his 2008 paper, I will focus my summary of his interpretation using his 2011 work alone.

¹¹⁹ I must leave aside the question of how Aristotle understands this concept of desire; Charles himself does not treat it as an important point of his argument. However the concept of ‘desire’ is understood, Charles appears to assume that it is a kind of antecedent to the movement of the animal. There is, unfortunately, not enough discussion of this issue in Charles’s paper, but it can certainly be criticized on the grounds that the desire to act seems to be conceived as a separate or distinct event that causes the bodily movements of the agent. Since Chapter 1 covers this issue of causal antecedents in detail, I leave it to one side in the present discussion in order to focus more thoroughly on the novel implications of Charles’s psychophysical account.

passages from *De Anima*.¹²⁰ The first includes Aristotle's elaboration on his comment that 'desire moves the animal':

De Anima 3.10: 'The instrument by which desire moves the animal is a bodily one: this is why it must be investigated among the functions common to body and soul...' (433b 19-20).

Charles also cites *De An.* 1.1¹²¹:

De An. 1.1(a): 'It is clear that the emotions are enmattered formulae and so their definitions will be of the following form: to be angry is a process of this type of body or part or capacity of such a body caused in this way for the sake of such and such a goal' (403a 24-7).

De An. 1.1(b): 'To be angry is a given type of process, the boiling of the blood around the heart for the sake of revenge' (403a 31).

De An. 1.1(c): 'No one considers the properties of matter which are inseparable [from this type of body/matter] not that is as properties separable [from this type of body/matter] but the physicist considers all the deeds and properties of this type of body and matter of this type...' (403b 10-12).

Charles concludes that an emotion, such as fear, anger, or desire, 'is essentially enmattered because its form is an enmattered form: one which is to be understood as essentially enmattered in this type of physical process' (2011: 79).

¹²⁰ All cited translations of these passages are Charles'.

¹²¹ Aristotle's remarks on the affections could also be added here: 'In most cases it seems that none of the affections, whether active [*πασχεῖν*] or passive [*ποιεῖν*], can exist apart from the body. This applies to anger, courage, desire and sensation generally, though possibly thinking is an exception. But if this too is a kind of imagination...even this cannot exist apart from the body' (403a 6-7).

As for Charles's further clause that desire 'resists decomposition into distinct physical and psychological components', he cites as support the following:

De An. 1.1(d): 'We have said that the affections of the soul are inseparable from the physical matter of living beings in the way in which anger and fear are inseparable and not in the way in which line and plane are' (403b 17-19).

Charles draws especially from the latter two passages, *De An.* 1.1(c) and *De An.* 1.1(d), to develop his understanding of desire as 'inseparable in definition as well as in existence from states of the body'. Aristotle's mention of 'the physicist' is meant, Charles argues, to draw a distinction from a 'mathematical' approach. For Aristotle, the study of 'physics' was expansive: it included natural phenomena, such as living organisms, and topics under the contemporary heading of the psychological sciences. Aristotle's mathematician, on the other hand, was concerned only with objects abstracted and separated from their perceptual reality. For this reason, Aristotle (in *De An.* 1.1(d)) contrasts anger and fear as inseparable from the 'physical matter of living beings' with certain separable objects such as a line and a plane. The emotions, being 'essentially enmattered', are substantially different objects of study and understanding from such geometrical objects. Charles characterizes the contrast between them in the following comparison:

anger and fear: they are existentially inseparable and inseparable in definition from perceptual matter

mathematical objects: they are existentially inseparable from but separable in definition from perceptual matter (2011: 77).

Anger, fear, and desire are phenomena that fall under the purview of Aristotle's physicist. They do not fall to the mathematician, because in abstracting and separating objects (in the case of the emotions, the properties of being angry, fearful, and

desirous) from their physical matter, he commits error ‘in his purely mathematical reasoning’.¹²² Charles’s argument for this point is that the mathematician who endeavors to abstract the psychological features that are essential to being angry, for instance, from the phenomenon of being angry (a process with physical properties), omits an important feature of the phenomenon itself. One who attempts to treat the psychological aspects of anger in abstraction from anger’s corporeal aspects may believe that his reasoning, qua mathematician, as to the subject matter is correct: for a mathematical subject matter, such as a line, remains the same irrespective of the corporeal nature in which it might be instantiated. The crucial omission that Charles points out arises from the mathematician’s egregious assumption that the psychological aspect of anger (i.e., desire for revenge) is similar to abstracted lines. Because they are not alike – for certain things are simply not mathematically abstractable from their corporeal realities – the mathematician who treats the psychological aspects as if they were abstractable without omitting essential information, misrepresents the phenomenon of anger from the start of his analysis. Thus Charles argues, ‘the body’s state is an important part of what accounts for the person’s being angry (or afraid). Omit reference to it and one fails to state the conditions under which someone is (e.g.) angry’ (2011: 78). Omission of the mathematical kind, i.e., of a kind parallel to what the mathematician does when studying geometric forms, etc., results in a failure of properly identifying the process to which both psychological and physical features belong. This is the error of ‘purely mathematical reasoning’ with regard to the affections.

The distinction between Aristotle’s physicist (and any kind of natural scientist¹²³) and mathematician is meant to bring forth Charles’s concept of a psychophysical process, as opposed to the abstracted and separated objects of mathematics. Charles explains:

¹²² Charles cites *Meta.* 13.3, 1078a 17; and *Phys.* 2.1, 193b 34-5, on this point.

¹²³ For Aristotle, the branch of physics includes the study of psychology.

The physicist, unlike the mathematician, studies the inseparable properties of this type of body as inseparable properties of this type of body, designed to play the role they do in this type of matter. If so, in studying the type of boiling of the blood involved in being angry, the physicist will study precisely that: the type of boiling of the blood whose goal is revenge. He is not concerned with a purely physical description of this type of blood boiling, which does not refer to the specific goals of the organism in question. In his view, there is no other process essentially involved in being angry other than the type of boiling of the blood which is directed towards revenge. The only process at issue is: boiling of blood for the sake of revenge (2011: 79).

One might think that Charles's explanation seems to imply that the concept of a psychophysical process is dependent on the physicist's interest in just such a concept, an interest that may appear arbitrary. Charles's claim is stronger than this latter suggestion, however. He wishes to establish that for Aristotle there *are* psychophysical processes, i.e., processes in which psychological phenomena are inseparable from the body's physiological occurrences, and that this inseparability is true in definition as well as in existence. These are processes for which one's definition — one that is to be used in scientific explanation — must refer to both the psychological and the physical features: if one of these is omitted then one fails to properly identify the phenomenon itself (even if one has identified the physical or the psychological feature that partially constitutes the phenomenon). Thus the definitional inextricability between psychological and physical features is one that, according to Aristotle, the scientist must recognize. If Charles is right, then the physicist's (or natural scientist's) interest is thus led by the fact that such psychophysical processes do occur in nature. Charles focuses on three examples of such occurrences, which will be considered next: Aristotle's case of the 'snub', his definition of a house, and the action of weaving.

4.2 Examples of psychophysicality

In *Physics* II.2, Aristotle defines the ‘snub’ as concavity in the nose (194a 5-8). The ‘snub’, argues Charles, is an example of the kind of phenomenon that falls under the purview of the physicist, for the ‘snub’ is an instantiation of a physical form that is more than mere concavity.¹²⁴ That is, if the ‘snub’ could be analysed in terms of mere concavity, or supposing that concavity could be abstracted and separated from the nose that possesses such a feature, then the ‘snub’ would be of interest to the mathematician, not the physicist. On Charles’s interpretation, however, ‘one cannot define the [physical] form correctly in terms of concavity, taken as the form of the nose. For to define the type of concavity of which we speak we have to refer to concavity-of-the-nose. ... Reference to the nose is required to make determinate the type of concavity at issue. It might be described as the determinant that specifies the determinate type of concavity which snubness is’ (2011: 79). The concept of concavity in which the mathematician is interested is therefore a determinable concept, i.e., an unspecified, abstract type of concavity. The physicist’s concept of the ‘snub’, understood as concavity in the nose, specifies the determinate type of concavity by referring to the physical matter of the nose. The ‘snub’, properly understood as a physicist’s concept, therefore can be conceived as a type of ‘physical form’, whose relation to its matter (of the nose) is very different from the relation that mathematical forms have to their matter. Although mathematical forms exist only as instantiated in physical matter (the number three exists only in the objects it instantiates or describes, e.g., three oranges or three days), the definition of mathematical forms is always in terms rendering them separate from physical matter (e.g., the number three is defined as the sum of one and two, or one more than two). Physical forms, such as the snub, ‘cannot be defined in this way: they have to be defined in terms which refer

¹²⁴ Although his successors (as here) very earnestly discuss Aristotle’s example of ‘the snub’, it should be noted that the latter serves merely as a playful example for him, i.e., one that may have exemplified a bit of self-deprecating humor.

to change and so invoke matter. Such forms are enmattered not simply in the sense that they have to be enmattered in (e.g.) physical lines to exist. Rather, they are enmattered in a far stronger sense: they are forms-in-matter, forms that involve matter to be the forms they are' (2011: 80).¹²⁵

Another example of such a 'form-in-matter' is Aristotle's definition of a house. One definition (the dialectical approach) points to the function or purpose of a house: it is 'a covering designed to prevent damage from wind and rain' (403b 4-6). This dialectical definition is insufficient for distinguishing a house from other structures that serve the same purpose, e.g., caves, tents, awnings, etc. It is necessary, therefore, to invoke the physical matter that uniquely constitutes the house, and that differentiates it from other structures that also 'are designed to prevent damage from wind and rain'. Charles explains: 'What makes the relevant covering a house is that it is one which is made from matter such as bricks and wood. While there are other types of covering, a house is a specific type of matter-involving covering' (2011: 80). Just as the 'snub' refers to the nose and thus makes determinate the type of concavity in which the physicist is interested, reference to the physical matter out of which a house is constituted is the necessary determinant that specifies the determinate type of covering that a house is.

Aristotle's concept of the 'snub' and his physicist's definition of a house are both, according to Charles, examples of the kind of enmattered form that includes desire, which moves the animal to act. The type of desire that moves the animal to act is 'essentially enmattered' in a specific type of physical process: presumably the latter can be identified as involving the animal's body, or certain physiological processes without which the desire to move would not be what it is (or could not be identified as such). This idea will need to be considered more extensively later. The present point

¹²⁵ It may be useful to note that, as used here, 'forms' have approximately the same meaning as 'kinds'.

is that the ‘snub’ and the house are comparable to the case of desire (or anger or fear), in that these phenomena are all ‘essentially enmattered’, both in existence and in definition, where the physical matter that constitutes each is specific to the particular thing or phenomenon itself. They are ‘enmattered forms’ in that the explanations of their natures require explanation of the particular kind of matter in which they occur.

The considerations so far are meant to explicate the concept of desire as a psychophysical process. They lead naturally to the concept of action as being understood similarly. Charles uses Aristotle’s example of weaving in *De An.* I.4, 408b 11-13. Weaving is an instance in which an agent intentionally moves her body ‘to achieve a given goal’ (2011: 81). It is a process that is characterized by both psychological features and physical properties, and ‘[t]he psychological features that are essential to weaving...are inseparable in definition from the processes with physical properties to which they belong. They are inextricably psychophysical features’ (2011: 81). In addition, ‘[t]he psychophysical features (specified in [the above]) are essential to the identity of the processes to which they belong: the processes to which the psychophysical features belong are essentially psychophysical processes (i.e. weaving). ... There is no other process (other than the one specified [as psychophysical]) which is essential to weaving’ (2011: 81). Charles argues that the foregoing definition of action is supported by Aristotle’s remark in *De Generatione Animalium*, that ‘the movements of the instruments employed [by the craftsman] contain the formula [*logos*] of the craft’ (744b 32 ff). Aristotle’s example is a metal worker who produces a sword, but the earlier example of the weaver applies as well. Charles comments that both the metal worker and the weaver are clear examples of the phenomenon expressed in Aristotle’s statement, that the ‘formula of the craft’ is contained within the movements of the craftsman. For,

his movements contain the formula (*logos*) of the art in that one is not able to define them (say what these processes are) without reference to

the goals of the craftsman and his/her skill. While they are essentially spatial, there is no defining what they are without reference to the goals and know-how of the craftsman. The relevant movements are the ones which are guided in the appropriate way by his/her skill. In the earlier case, the weaver spins, cards the wool, and ties the knots in the way required to make the coat she aims to produce. Her actions are vivid examples of inextricably psychophysical processes (2011: 87).¹²⁶

One can note further that the psychophysical analysis of a weaving machine, a corporeal object by which the weaver spins, includes the concept that such a thing is an instrument for producing thread. Aristotle's definition in *De An.* I of 'soul' is that it is the first actuality of a living *organic* body, where the sense of 'organon' corresponds closely to 'tool' or 'instrument': one that is naturally used for realizing the animal's vital activities. Unlike a weaving machine or a bicycle, an animal's body is a corporeal instrument having several specific purposes or functions, and the psychological aspects that are needed to implement some of these purposes (i.e., the agential) are even more essentially tied to the corporeal instrument that the animal 'uses'.

¹²⁶ Elsewhere, Charles emphasizes that his understanding of action as an 'inextricably psychophysical process' is based on the way in which explanation of the elements of action refer to one another, rather than a mere *a priori* claim about the craftsman's skill being 'necessarily enmattered' in certain physical processes. Furthermore, the term 'inextricably psychophysical' goes beyond merely combining psychological and physical elements of action. Charles writes: 'Activities (or processes) which are "common" to the body and the soul, as we have seen, cannot be defined as a combination of psychological and physical components. One cannot say what they are in purely psychological terms as they are types of expanding and contracting, becoming hard and soft, which generate pushing/pulling. Like anger, they are not simply necessarily enmattered in some physical process. Rather, one cannot define what they are without reference to expanding or contracting, pulling or pushing, and so forth. The latter features are of their very essence. Nor can one define the relevant type of physical process without reference to their psychological aspects. This is what I intend by calling these processes "inextricably psychophysical"' (2011: 86-87).

4.3 Aristotle's contemporary contribution

There are two important features that are worthy of mention, concerning Charles's account of action as psychophysical process in general. The first is noteworthy because it expresses a central aim of Charles's account, which the latter arguably achieves. This first issue relates to the distinctive position that Aristotle's approach occupies within the contemporary landscape presumed by most philosophers of mind and action. Aristotle's view is distinctive but also appears unsatisfying, Charles notes, 'because it is not one of the familiar options of post-Cartesian philosophy (dualism, materialism, functionalism, or spiritualism). It should be seen rather as offering a radical alternative to these traditional accounts, challenging the basic assumptions that drive them' (2011: 76). Although dualists, materialists, functionalists, and spiritualists may be surprised to find themselves grouped together as such, Charles identifies two propositions that, according to his argument, are common assumptions normally made by 'post-Cartesian' theories. In the case of anger, where one experiences a desire for revenge, the post-Cartesian approach (inclusive of dualists, materialists, functionalists, and spiritualists) normally assumes the following two propositions (2011: 76):

pC[1]: There is a purely psychological feature or process involved in desiring (or a purely psychological description), such as desiring (or aiming at) revenge.

pC[2]: All relevant processes and features essentially involved in desiring are either purely psychological or purely physical or a combination of the two.

As the foregoing discussion on the 'inextricably psychophysical' already suggests, Charles's primary aim is to show that Aristotle rejected the idea of any 'purely psychological' or 'purely physical' feature or process as being involved in his concept of action and its causal antecedents (in the form of desiring to act). Aristotle's

understanding, according to Charles's interpretation, was one that rejected the notion of 'purely' physical or psychological features or processes not only in existence but also in definition, and he rejected the possibility that action might be thought of as a combination or aggregate of such features or processes.

Charles argues that these post-Cartesian assumptions, pC[1] and pC[2], play a critical role in the predominant approaches to interpreting Aristotle's thought on the matter of desire. Perhaps the most common is a 'broadly materialist' interpretation.

According to the materialist,

M[1]: Desire is to be defined as the psychological process it is without reference to any physical phenomena (in purely psychological terms).

M[2]: Desire is realized in (or constituted by/or supervenes on) a particular process of a physical/material type.

M[3]: The relevant type of physical/material process can be defined without reference to any psychological state or description (2011: 76).

A far less popular alternative is dualism. Charles supposes that '[d]ualists would present desire as the purely psychological efficient cause of the material change involved in action. They would also see desire as the manifestation of a capacity of a separate substance: the dualist's [concept of the] soul. [The latter] moves the body by somehow interacting with the body (e.g. in the pineal gland)' (2011: 76 fn. 1). The obvious problem for the dualist's approach is that it violates the principle of causal closure, i.e., the thought that physical effects (such as a movement of the body) must have physical causes. But even if there is a physical cause, as the broadly materialist approach claims in M[2], the presumption of pC[1], i.e., that there must be a 'purely psychological' process or feature in the desiring that moves the body, combined with causal closure appears to render the 'purely psychological' feature or process causally inefficacious. According to (Charles's interpretation of) the materialist, desire is a

psychological process or feature defined without reference to any physical phenomena, although it is realized in or determined by the physical process of which the ensuing bodily movement is an effect. It is not clear whether Charles's analysis distinguishes between reductive and non-reductive materialist interpretations of Aristotle on this point.¹²⁷ For the sake of Charles's present point, however, it possibly matters not whether desire is 'reduced to' (identified with) the underlying physical cause of the bodily movement, or is understood as realized in these physical phenomena.¹²⁸ For both, according to Charles, presume that the psychological event of desiring is separate in definition from the physical occurrences in which the 'purely' psychological event is realized or with which the latter is identified. It is because pC[1] and pC[2] are presumed that either variant of materialism faces a causal exclusion problem, in which the psychological feature that is believed to be important to the account of action is in fact excluded as an inefficacious or irrelevant feature of the account.

Charles's primary aim is merely to show that Aristotle provides a wholly distinctive alternative to the contemporary approaches (out of which only a select few have been

¹²⁷ In Charles's 'Aristotle's Psychological Theory' (2008), not discussed in this chapter as it is an earlier version of his 2011, Charles does mention that non-reductive, functionalist interpreters of Aristotle include Hilary Putnam and Martha Nussbaum (1995), whereas non-functionalist and non-reductive materialist interpreters include Charles's younger self (1984). A spiritualist interpretation, not discussed in this chapter, is exemplified by Myles Burnyeat (2002, 1993, 1992) and Thomas Johansen (1998). See Charles (2008) p. 1, fn. 2.

¹²⁸ A question arises at this point concerning the compatibility of the reductive materialist's account of desire as being identical with neurophysiological events and Charles's analysis of the materialist as assuming that desire 'is to be defined as the psychological process it is without reference to any physical phenomena (in purely psychological terms)' (2011: 76). In other words, there is a question about whether the stipulation pC[1] in particular is truly and critically an assumption held by the reductive materialist, as Charles claims it is (with pC[2], which may be less controversial), given that the materialist believes that psychological phenomena are thoroughly reduced or realized by physical (neurophysiological) events. I will address this question in the succeeding section.

summarized here). He does not identify the causal exclusion problem as specifically the obstacle to overcome. However, I believe the question is one worth considering because it seems genuinely to arise from acceptance of the stipulations $pC[1]$ and $pC[2]$, and because *if* Charles is right that Aristotle provides a radically different, yet simultaneously intelligible alternative to the contemporary approaches, then it also seems true that his account is one that successfully avoids these legitimate problems (of both causal exclusion and causal closure). The latter would count as a real achievement of Aristotle's account of action as psychophysical process, one that is noteworthy on its own terms as well as located within the contemporary, 'post-Cartesian' landscape presumed by action and mind, at least as Charles characterizes it.

4.4 Prevention of disappearing agents

The second purported achievement of the account of action as psychophysical process is one specifically mentioned by Charles himself. One interpretation of Aristotle's account of desire as moving the body is that the parts of the body are moved by the soul, or passions of the soul. Charles cites Menn (2002) as proposing such a view. According to Menn, 'the soul, as user of the body, is separate from the body, standing to it as user to used (as the pilot stands to the rudder he uses to steer the ship). Further, in his view, since the soul is separate (in this way) from the body, it is not itself bodily but must be non-bodily' (2011: 91). Charles argues that Menn's interpretation is but a mere step away from Cartesian dualism, for there may be aspects of Menn's account that are not incompatible with the account of action as psychophysical process.¹²⁹ In any case, Charles clarifies that the account of action as psychophysical process is one in which it is more accurate to speak of 'the person (or

¹²⁹ For instance, Menn's interpretation suggests that a hand is a merely bodily instrument that is used by the soul, which moves it as something separate and distinct from it (2011: 92). But Charles responds that '[i]t does not follow from the fact that the soul uses separate instruments that its using them is anything other than a psychophysical process, essentially the realization of a hylomorphic compound (e.g. the faculty for desire)' (2011: 92).

the composite), not the soul, that moves his hand, grieves, and so on', for 'the composite, ... not the soul, is the appropriate subject of the essentially psychophysical processes that lead to action' (2011: 92).¹³⁰ Another achievement of Charles's account, as he claims it, is that it manages to retain the sense in which action is something belonging to an agent, rather than one of the agent's parts (e.g. in this case, the soul). He explains:

For the soul, conceived of as a purely psychological phenomenon (or set of such phenomena), is not the type of thing which can be the subject of these inextricably psychophysical processes. At this point, Aristotle is best seen as challenging the moves that lead to dualism at the first step: it is not the soul (properly speaking) that moves the animal but the (psychophysical) person (2011: 92).

On Charles's view, a further implication of the foregoing argument can be usefully drawn out in order to retain the sense that action belongs to an agent, and not one of her parts. The 'moves that lead to dualism', as illustrated by Menn's suggestion, potentially lead to a problem reviewed earlier in the last chapter, one that is identified in action theory as the 'problem of the disappearing agent'. In the case of a dualistic approach such as Menn's, the problem is represented as centering on a critical part or feature of the agent as being the causal locus from which bodily movement is effected. The result is that the intuitive conception of the human being as possessing efficacy with respect to her own actions is diminished, or wholly 'disappears'. If bodily movement is produced by the soul, not the whole human being, then the sense in which actions belong to a human being qua agent – so it is argued – begins to be lost among the apparent power of her parts.¹³¹

¹³⁰ This clarification is based on Aristotle's point to the same effect, in *De An.* I.

¹³¹ I should note that, as in earlier parts of this thesis, I do not quite see the force of the disappearing agent problem. In this case, the prospect of a disappearing agent seems

The possibility of this problem arising is not, in addition, unique to the ‘moves that lead to dualism’: the ‘disappearing agent’ problem threatens wherever there is identification of a causal locus for action in an agent’s parts rather than in the agent herself. Varieties of materialism therefore should be regarded as facing the same objection. The general idea is that the physical realizers of an agent’s desiring to move her body are what cause the physical movements to come about. But the physical realizers of an agent’s desiring to move will always represent processes at a sub-agential level, i.e., neurophysiological occurrences that exist entirely outside the awareness of the agent herself.

Aristotle’s account of action as psychophysical process, according to Charles, avoids the problem of a ‘disappearing agent’ by properly locating the causal nexus between desiring and bodily movement in the animal agent, which is itself a psychophysical thing. Although Charles does not explicitly acknowledge the connection, it seems obvious that it is only in virtue of the animal agent’s being a psychophysical thing that both desiring and intentionally moving one’s body may be considered instances of a psychophysical process. If Charles were to define the animal agent as a psychophysical thing, plausibly it would follow the same structure already attributed to the concepts of desire and action: i.e., that an animal is the kind of thing possessed simultaneously of both physical and psychological properties or parts, which exist not in combination as ‘purely physical’ or ‘purely psychological’ but rather ‘essentially inextricable’ in Charles’s sense. That is, the physical and psychological features or parts of the animal are inextricable both in existence (so long as the animal is alive) and in definition, such that its psychological features (e.g., of perception or thought) cannot be

even less of a concern given a proper rejection of the notion that the soul is merely one of an embodied being’s parts, possibly among several. For Aristotle, the soul is the source of action; it exists as embodied, and an embodied soul (where conditions of rationality are met) just is, for Aristotle, the human being. In the cited passage, Charles’s term ‘person’, in addition, is a peculiar choice given that Aristotle’s account contains no correlative concept.

accurately defined without reference to its physical features, and its physical features (e.g., its brain states or physiological responses throughout its engaging in activities of perception and thought) cannot be accurately defined without reference to its psychological features.¹³²

4.5 Difficulties in developing the account

Charles's account raises a number of interesting issues, not least of which (with regard to the stated purposes of this chapter and the thesis in general) are questions concerning (i) the intelligibility of the account as providing a potential theory of action, and (ii) in particular, the adequacy of the account for developing a plausible disjunctivist view of action. Although Charles does not countenance the possibility that his account of action as psychophysical process may be appropriated for the development of a disjunctivist approach, it is not difficult to see how his concept might serve as one such inspiration. According to Charles's interpretation, Aristotle's account of action as a psychophysical occurrence is analysed in terms of psychological and physical processes being 'essentially inextricable', both in existence and in definition. What he means by this, as the preceding section suggested, is that certain physical (or physiological) events or processes involved in action cannot be defined

¹³² Obviously it is false that certain other physical features of the animal, e.g. its digestive or circulatory functions, cannot be defined without reference to psychological features. (I exclude from this observation the fact that plants, which lack any rudimentary form of sense perception, nevertheless are permitted 'nutritive soul' by Aristotle. If they were to be included in the psychophysical account, then we would subsequently need to revise our modern concept of 'psychological', which is assumed to be somewhat interchangeable with 'mental'.) They can and indeed must be defined without such reference, because that accords with the nature of the processes that they are: i.e., purely physiological processes that occur for the most part outside of the awareness or agential control of the animal. This seems only to further support the main point of Charles's account of action, which claims that a concept of the agential is intelligible only wherever 'essentially inextricable' psychophysicality occurs. The animal as an agent is therefore legitimately defined as the type of thing whose psychophysical features or parts define it in its essential role, i.e., as an agent or 'mover'.

without reference to the psychological events or processes that characterize the moments throughout which an agent desires to move and carries out her desired task. According to Charles's account, therefore, wherever there is action, the constitutive events, parts, and processes of the action cannot be 'purely' physical or 'purely' psychological, for they are 'essentially inextricable' from one another in just the way defined. In contrast, a mere movement – an involuntary response or non-voluntary process of the body – lacks the 'essentially inextricable' feature of action.

This is not to say, however, that there cannot be both physical and psychological occurrences somehow involved in an involuntary (or non-voluntary) movement: but where these do occur simultaneously in such a way, they are related to one another in a merely aggregative combination of 'purely' physical and 'purely' psychological processes, i.e., the very type of combination that Charles denies can ever be characteristic of the concept of action. Consider, for example, the application of this idea to Davidson's case of the climber who loosens his hold on the rope (to which his unfortunate companion is attached).¹³³ In letting go of the rope, the climber fails to act at all despite having the kind of desire and corresponding belief that the causalist claims are necessary and sufficient components in the causation of action. Where Analogy 1 in the last chapter considered how assuming a causalist structure made it difficult to find a solution to this problem of deviant causal chains, the disjunctivist approach to agency based on Aristotle's concept of action as psychophysical process is, by comparison,¹³⁴ able to explain why 'letting go of the rope' in this case happens to be a mere movement rather than an action, despite the fact that the climber (according to Davidson's example) possessed all the requisite antecedent psychological elements to produce an action. The disjunctivist explanation claims that 'letting go of the rope' under the circumstances is a physical movement consequent to the

¹³³ See §1.1 of Chapter 1 for a full description of Davidson's example.

¹³⁴ See Chapter 1 for a discussion on the causal theory's capacity for satisfactorily handling the problem of deviant causal chains.

occurrence of psychological states (i.e., the climber's desire to lessen the weight of his burden and corresponding belief that if he loosens his hold on the rope his burden will be lessened) that cause a further involuntary physiological state (of trembling). The consequent physical movement of loosening his hold on the rope is therefore not a physical process inextricably tied to the relevant psychological events that happen to precede the climber's loosening his hold on the rope. For the type of physical movement that is 'letting go of the rope', in this case, is the type of mere bodily movement that *could* have occurred in exactly the same way under very different circumstances: for instance, the climber might have been knocked unconscious by a falling rock, thus involuntarily releasing his hold on the rope (himself having already been secured by climbing hooks). That is, the very same type of physical movement may occur (as instantiated in what happens) had the particular psychological antecedents cited in the original case been absent. The climber's desire to lessen the weight of his burden and his corresponding belief that loosening his hold on the rope will lessen the weight of his burden are, in this case, entirely accidental (and coincidental) with respect to the ensuing event of his loosening his hold on the rope. 'Letting go of the rope' in this case is not the kind of movement in which the physical happening must be inextricably connected to any psychological element at all.

In contrast, the disjunctivist's claim, assuming the account of action suggested by Charles, is that the concept of action is of a fundamentally different type from that instantiated by Davidson's case of 'letting go of the rope'. Action is the kind of event or process that requires both psychological and physical elements, not in such combination that preserves their 'independence' from one another, but instead demands a psychophysical relation that is 'inextricable' both in existence and in definition.

4.6 Contesting the contemporary alternative

There are many questions that arise from this disjunctivist interpretation of action as a psychophysical process, which will be addressed in the remainder of this chapter.

From a contemporary standpoint, perhaps the most objectionable feature of Charles's account in general is his characterization of the materialist interpretation of an agent's desire to act. Charles claims that 'spiritualists'¹³⁵, substance dualists, and materialists alike necessarily assume pC[1] and pC[2], which may be re-stated as requiring that desiring to act involves a 'purely' psychological feature or process, i.e., where 'purely' psychological indicates that desiring to act is separate from any physical feature or process, both in existence and in definition (pC[1]); and (pC[2]) that 'all relevant processes and features essentially involved in desiring are either purely psychological or purely physical or a combination of the two' (2011: 76). But it is doubtful that most materialists will agree to these claims: in particular, they may find pC[1] objectionable. Some reductive materialists, for instance, are eliminativists with regards to psychological phenomena. Such claim that the identification of such psychological phenomena with their physical realizers entails that the conception of an action's having causal antecedents that are specifically psychological is conceptually unnecessary. In other words, psychological concepts such as desiring and intending may be wholly eliminated given that such phenomena have an absolute grounding in their physical base. Materialists of this kind will therefore deny that desiring to act is

¹³⁵ According to Charles, spiritualism is the view that desire is a 'purely' psychological process, not realized in any physical process at all 'although it may require the presence of certain physical necessary conditions' (2011: 76). One should note also that spiritualism is merely a hypothetical position that emerges from Burnyeat's discussion of Aristotle on sense perception, although it is not a position that Burnyeat himself nor anyone to my knowledge holds (1992). It is unclear how this position distinguishes itself from substance dualism or even non-reductive materialism, and Charles does not elaborate. For the sake of simplicity, in this discussion I will consider only the commonly considered alternatives to reductive materialism: non-reductive materialism, and (to a less common extent) substance dualism.

(or even merely involves) a ‘purely psychological’ feature or process, since they claim that such a definition of desire should be eliminated from the account entirely.

Non-reductive materialists may also find Charles’s characterization of ‘post-Cartesian’ philosophy of mind and action objectionable. A non-reductive account assuming some variant of supervenience may conceivably also reject pC[1] and pC[2], for one could assume a conception of supervenient relations between the psychological on the physical that requires the latter to be defined with reference to the other, and vice versa. I explore this specific possibility in greater detail in Chapter 3, and thus leave aside its particular formulation until then. The point at present is merely that Charles’s characterization of ‘post-Cartesian’ philosophy of mind and action does not appear to be a wholly fair characterization across the varied accounts to which he claims it applies. Indeed, it is difficult to see where, among the contemporary alternatives proposed (i.e., spiritualism, functionalism, and materialism), both pC[1] and pC[2] necessarily are assumed, with the obvious exception of substance dualism. Charles’s description of a ‘post-Cartesian’ philosophy of mind and action may certainly be thought to apply accurately to varieties of substance dualism, but the latter view has comparatively few subscribers.

This potential drawback of Charles’s analysis does not, however, necessarily impinge on the account of action that he offers as an alternative to the ‘post-Cartesian’ theories cited above. Although Aristotle’s concept of action as a psychophysical process may not ultimately constitute as radical an alternative (in particular, to non-reductive materialism) as Charles imagines, his psychophysical concept may yet yield an intelligible theory of action of independent interest to action theorists, regardless whether such a concept is thought to be compatible with certain ‘post-Cartesian’ approaches, such as non-reductive materialism.¹³⁶ Nonetheless, given that Charles’s

¹³⁶ Its potential compatibility, indeed, may count in its favor in contexts where a non-reductive materialist approach is appealing, or merely in accounts where

main point of opposition throughout his discussion *is* the contemporary standpoint with which he finds fault, a thorough evaluation of his account necessarily includes the consideration of the latter's merits relative to the 'post-Cartesian' approaches that he opposes. In what follows, however, the discussion focuses on the concept of action as psychophysical process, largely considered apart from Charles's critique of 'post-Cartesian' philosophy.

4.7 Non-decomposability

A central issue in the Aristotelian account reviewed concerns the claim of non-decomposability. Charles claims that the desire to act (or some other comparable psychological antecedent to action) and the action itself both exemplify a type of phenomenon that 'resists decomposition into distinct physical and psychological components' (2011: 77). It is apparent that Charles views his account as resisting decomposition not only 'in existence' but also 'in definition'. The question whether his account succeeds in both of these respects certainly merits examination, but there is also a preliminary question concerning the intelligibility of the distinction between 'in existence' and 'in definition'. Consider for instance Charles's discussion of the 'snub'. He describes the snub as a psychophysical concept, because it is an instantiation of a kind of physical form that is of interest to the physicist rather than the mathematician: 'Such forms are enmattered not simply in the sense that they have to be enmattered in (e.g.) physical lines to exist. Rather, they are enmattered in a far stronger sense: they are forms-in-matter, forms that involve matter to be the forms they are' (2011: 80). It seems clear from Charles's account that the claim of inextricability 'in existence' is meant to amount to more than merely stating that the snub cannot exist apart from this particular nose, and this particular nose (which is posited to be a snub nose) cannot exist apart from the snub, or 'concavity-in-the-

supervenience of the mental on the physical is presumed. The suitability of pairing a psychophysical account of action with a supervenience approach receives due consideration in the succeeding chapter.

nose'. Rather, inextricability 'in existence' involves the claim that the snub is the sort of thing that requires a particular nose in order for the snub to be the form that it is: i.e., a physical concavity in the nose. An important implication of this concept of inextricability 'in existence' is that it excludes the idea of a 'substrate nose', i.e., nothing but 'nose-flesh' without any additional features, a physical template to which concavity is added to create a snub nose.

It also seems clear enough from Charles's account how inextricability 'in definition' is distinguished from inextricability 'in existence'. Definitional inextricability entails that the snub must be defined including reference to the nose, and the snub nose defined with reference to the feature of 'concavity-in-the-nose' that it possesses. Although they are sufficiently distinguished, there remains a question concerning how inextricability 'in definition' and 'in existence' are related. Charles passes quickly over this issue without remark,¹³⁷ perhaps because it may simply seem *prima facie* obvious that definitional inextricability depends on psychophysical inextricability 'in existence'. One may think that if the latter is assured, it is safe to assume that inextricability 'in definition' is likewise true. But the previous consideration of mathematical forms suggested a connection between definition and existence very much to the contrary: numbers are necessarily enmattered 'in existence', but their inextricability from matter 'in definition' is not likewise entailed: for 'in all such cases mathematical forms are to be defined independently of all physical matter, even though they are realised in physical matter' (2011: 80). Thus more needs to be said in support of the claim that physical forms such as desire and the snub are inextricable from matter 'in definition', since it is evidently not on the basis of metaphysical inextricability (that is, 'in existence') alone that their inextricability 'in definition' is assured.

¹³⁷ Indeed, Charles hardly discusses the concept of inextricability 'in existence' at all. He only ever mentions it in somewhat trivial conjunction with inextricability 'in definition' which takes center stage in his discussion.

There may indeed be doubt over whether metaphysical inextricability is really ever questioned in the first place, and thus whether its place in Charles's account involves more than a trivial part. This question appropriately arises from a contemporary standpoint, whose alleged core assumptions are Charles's primary target. A contemporary philosopher of mind or action may find the psychophysicalist's insistence on inextricability 'in existence', as standing in opposition to 'post-Cartesian philosophy', somewhat puzzling: unless he or she happens to be one of the rare proponents of substance dualism. But most materialists would likely have no objection to the idea of inextricability 'in existence', e.g., that a craftsman's know-how and goals¹³⁸ in exercising his skill cannot be separated from the material elements and processes involved in such skill, at the moment that it is being exercised. For instance, the non-reductive materialist may understand the craftsman's know-how and goals as supervening on the material elements and processes involved (i.e., the craftsman's hands and fingers, the tools with which he works, the wood from which he fashions his product); it is possible (as mentioned earlier) to understand the supervening relation between these processes as requiring the type of inseparability 'in existence' that has been described.¹³⁹

Inextricability 'in definition', on the other hand, may not enjoy so easy an appeal. For materialism in general offers a reductive approach, one that strives to understand the whole by analysing each of its parts separately, such that requiring the definition of

¹³⁸ The terms 'intention' or 'desire' might be substituted here, as it seems to make no difference to the point that Charles makes.

¹³⁹ A proponent of psychophysicalism (as Charles has presented the view) may object that the non-reductive materialist in this case fails to appreciate the deeper point concerning inextricability 'in existence' (and, by extension, 'in definition'): i.e., that the form is necessarily enmattered for it to be the form that it is. Still, it is not inconceivable that a supervenience theorist could understand the relation between psychological and physiological processes in such a way as to be compatible even with this deeper point. The analysis of top-down causation suggested in the succeeding chapter, for instance, plausibly presents one such conception of form that cannot be what it is without the necessary and integral involvement of matter.

one part to be in terms of the other would seem counter-productive to its aim. Even non-reductive materialism commonly treats physical and psychological processes as if they respectively occupy separate metaphysical ‘levels’ (see Chapter 3 for discussion of this characterization). As before, it is unclear what should attract contemporary theorists to the notion of definitional inextricability between the psychological and physical processes or elements of action, for it obviously cannot be assumed along with psychophysical inextricability ‘in existence’. There is a further point to be added given these considerations, concerning the contemporary theorist’s denial of definitional inextricability. Charles argues that the rejection of inextricability ‘in definition’ entails the acceptance of pC[1] and pC[2], two ‘common assumptions’ of ‘post-Cartesian theories’ whose core claim involves ‘purely’ psychological features or processes and ‘purely’ physical features or processes’ (2011: 76). The preceding has summarized these assumptions as wholly excluding, in the definitions of the psychological or physical processes or features (e.g., involved in the craftsman’s exercise of his skill), any reference to one another at all. But merely opposing Charles’s concept of inextricability ‘in definition’ does not require that one therefore accept definitional *extricability*. It seems conceivable, and indeed very likely, that varieties of materialism would presume at least a loose association between the mental and the physical. Such an association need not be as strong as Aristotelian psychophysical inextricability, but it need not also be so weak as to be entirely absent, as Charles supposes in pC[1] and pC[2].

4.8 Inextricability ‘in definition’

If, as Charles claims, Aristotle’s account of action as a psychophysical process is to represent an intelligible and compelling alternative to the contemporary approaches in mind and action available at present, then the difficulties reviewed in the foregoing need to be resolved. I will not attempt to do so here, since it is clear that a good deal of further metaphysical analysis is needed to secure Charles’s argument for action as a

psychophysical process; that further analysis will be delayed until Chapter 3. The above discussion is not entirely in vain, however, for it has helped to clarify that the really contentious point in Charles's account, i.e., the point that is truly worth developing, is what he calls psychophysical inextricability 'in definition'. It is hopefully clear by now that it is the latter, and not inextricability 'in existence', which critically supports the central aim of the account: namely, to understand action as 'indecomposable', or as exemplifying the type of phenomenon that 'resists decomposition into distinct physical and psychological components' (2011: 77). Non-decomposability, it appears, relies on a plausible account of definitional inextricability, which admittedly faces a number of challenges. Nonetheless, successfully resisting the decomposition of action into distinct physical and psychological components will constitute an important feature of the Aristotelian account of action considered in this thesis, as distinguished from the standard causal theory reviewed in Chapter 1.

The causal theory proposes that an event is an action if and only if it is caused in the right way by a bodily event joined by a condition of mind (i.e., a mental event or state). Its crucial presumption is that the physical and the mental can be distinguished into different types of events with each of their respective causal efficacies. In other words, the causal theory accepts the decomposition of action into distinct physical and mental parts, which exercise distinct causal attributes. In contrast, the Aristotelian disjunctivist approach to action, which was introduced in Chapter 1, necessarily resists this causalist tendency. For on the disjunctivist account, the craftsman's knowledg-how, his intention and desires (and his goals), and other psychological processes as he exercises his craft, are not simply psychological component(s) that can be added to his bodily movements as if the latter constitute a physical substrate, to produce action. The mental or psychological processes involved in action can no more be added to an underlying 'template' physiological process to produce action, than the form of the

snub can be added to a bare nose to create *this* particular snub nose.¹⁴⁰ Instead, according to the disjunctivist approach, the respective physiological processes involved in action and mere movement (whether involuntary or non-voluntary) are wholly different. Their fundamental difference is conceivable on the disjunctivist account of action, but not according to the causal theory, whose assumption of decomposition entails that the physical component of action is the same common factor for mere movement as well.¹⁴¹

The concept of action as a psychophysical process, despite requiring further clarification and analysis, offers a plausible explanation for the fundamental physical difference between action and mere movement, as proposed by the disjunctivist account. Definitional inextricability claims that it is not possible to define the craftsman's state of mind (including his knowledge-how of the craft at hand) without reference to his bodily movements and the tools and matter that constitute his craft. This so far seems obvious, since the psychological elements and processes of a physical action (given that the inquiry of this thesis focuses only on physical actions) are, after all, elements and processes that necessarily involve something physical for them to be what they are, i.e., the psychological elements and processes of an action. But definitional inextricability also claims that it is not possible to define the craftsman's bodily movements or physiological processes without reference to the craftsman's state of mind (which includes his know-how and skill).¹⁴² There is a very limited interpretation based upon which even the causalist may accept this aspect of definitional inextricability to be true, since the causalist agrees that for an agent's

¹⁴⁰ Nor can a certain geometrical form be added to a triangle to produce an equilateral triangle, nor a specific melody added to a set of notes to create Beethoven's 7th Symphony.

¹⁴¹ See Chapter 1 for exposition of this claim.

¹⁴² This is the aspect of definitional inextricability that a materialist may find difficult to accept; for the reductive materialist, this claim would be impossible to affirm. But there may yet be space for a supervenience theorist to assume it, as the next chapter endeavors to show.

bodily events to constitute an action, reference must be made to the agent's psychological states or events, which are causally joined to the bodily events in the appropriate way.

There is a stronger reading of definitional inextricability, however, that adopts the Aristotelian insight (as interpreted by Charles) that a physical form, such as the snub, is not the form that it is without being understood as the kind of form that is enmattered. Similarly, the stronger interpretation of definitional inextricability claims that it is not possible to define or understand the physiological processes of the craftsman's body as he works, without reference to his know-how and other psychological states *qua* intrinsically matter-involving, because only in virtue of this latter quality are psychological states or processes able to exert a causal influence on what happens involving the craftsman's body. Indeed, only *qua* enmattered kinds do the psychological states and processes of the craftsman play a role, as he works, in conforming the physiological processes of the craftsman's body to his specific desires and ends. In consequence, it is not possible to understand his bodily movements as he works as the kind of bodily movements that constitute an action, without acknowledging that such bodily movements are of a kind that, by definition, is causally tied to the craftsman's state of mind in pursuing his craft. No doubt this latter type of explanation will be accused of circularity, for it appears that the definition of the type of movement that counts as an action is in terms of the type of movement understood as connected to certain agential features, i.e., the psychological events and processes characteristic of agency. As a number of philosophers have argued in the last few decades, however, the concept of action as something psychophysical may have expression in several contemporary variants, without falling into explanatory circularity.¹⁴³ Again, the claim of definitional inextricability (as clarifying action

¹⁴³ These contemporary articulations of action as something psychophysical include those offered by Fine, Jaworski, Lowe, Cohen, Hanna and Maiese, and Haldane.

understood as a psychophysical process) requires a deeper analysis, which is the concern of the succeeding chapter.

I will end with some concluding remarks for this chapter. Both Coope and Charles offer interpretations of Aristotle that, in different ways, offer attractive material for the contemporary disjunctivist about action. In both cases, however, what is offered to the contemporary disjunctivist involves certain Aristotelian ideas that may seem *prima facie* strange or pre-modern: i.e., in Coope's interpretation, this may be true of Aristotle's account of change as intrinsically end-directed, while in Charles's approach, the assumption of definitional psychophysical inextricability may seem odd. Philosophers who are not already sympathetic to Aristotelianism will complain that the price of purchasing disjunctivism, i.e., in the form of one of these pre-modern, strange ideas, is simply too high. My reply to this complaint is that it is not yet possible to determine such a conclusion. As I have already mentioned, one of the reasons for examining a serious rapprochement between Aristotle and contemporary discussions in this thesis is that this heretofore has not been done yet with sufficient clarity. Once Aristotle's thought is better understood, with the perspective of the contemporary lens, it will be possible to determine whether purchasing disjunctivism at a price of Aristotelian thought really is too high, or whether perhaps certain post-modern assumptions about science and the physical realm may need to be abandoned. In the following chapter, I attempt to bring the contemporary lens closer into focus on this question.

Chapter 3:

Contemporary Adaptations of Psychophysicality in Action

Skeptics of Aristotelian psychophysicality believe that there is no credible theory of action that can improve on causalism using the notion of hylomorphism. I consider a recent approach that argues to the contrary: Hanna and Maiese's account in *Embodied Minds in Action*, which offers a contemporary Aristotelian version of a 'non-classical causal theory of action' (2009: 157). Hanna and Maiese's argument for 'trying' in action has the compelling advantage of avoiding deviant causal chains, but some of their assumptions are problematic. Their proposal of property fusion holds great interest, but it remains unclear whether their overall model succeeds in distinguishing itself from other solutions, especially supervenience theories in general. In addition, the approach to action defended by proponents of 'trying' is not incompatible with a variant of agent causation recently proposed by Steward, and I suggest that a specific agent causal view that spells out animal agency as top-down control over the parts of the subvenient base, can be viewed as an extension of the psychophysical composite account inspired by Hanna and Maiese.

1.1 Introduction

In the previous chapter, I argued that there is a conceptual link between disjunctivist theories of action and psychophysicality, and that the connection between them can be clarified by understanding Aristotle's concept of action as a psychophysical process. It remains the case, however, that at its core, disjunctivism relies on a notion of psychophysical unity (and, in particular, definitional inextricability) that requires further explication. The purpose of this chapter is to try to resolve some of the questions left open at the conclusion of the last chapter. The central question of our inquiry is: does the metaphor of 'suffusion' between the psychological and the

physical (the ‘inner’ and the ‘outer’), as proposed by recent disjunctivists of action (Hornsby, Haddock and Macpherson, Hanna and Maiese) hold explanatory promise? How plausible are the contemporary expressions of psychophysicality in action? One recent attempt on this subject, Hanna and Maiese’s *Embodied Minds in Action* (2009), makes explicit the goal of revitalizing an Aristotelian psychophysicality, in order to establish the latter as a viable alternative in contemporary philosophy of mind. The attempt is not without problems, as may be expected given reasons cited in the last chapter: for Aristotle’s system of concepts does not correspond easily to the modern philosophical framework. Nonetheless, it is illustrative at least of the fact that a contemporary explication of psychophysicality in action is an immensely difficult enterprise. Acknowledgement of the latter can be useful as well, both for contemporary proponents of some variant of psychophysicality and for Aristotelians themselves.

One of the main aims of this final chapter is to argue for the thesis that action is a species of movement that is distinctive of a living animal. This thesis perhaps seems obvious, but it has been rejected in a surprising variety of ways. Rejection of the thesis may be summarized as adopting one of two general approaches. Some philosophers, along with the majority of scientists, claim that action consists in, no more and no less, the movement of molecular particles and quantum entities that make up the animal agent. These are ‘reductionists’ about action. Others deny that any animals other than human agents are able to perform actions, for reasons having to do with their conception of practical rationality. The latter group can be identified as ‘intellectualists’ about action. Despite their different approaches, both reductionists and intellectualists about action fall into a causalist category of understanding agency: they both assume that there is a common physical element shared by both actions and non-actions, i.e., a common physical element that either makes up the essential constituent of action (as is claimed by the reductionists about action) or is rejected as the distinguishing feature of action over mere movement (as is claimed by

intellectualists about action). Between the two general approaches, the notion of action as a causal relation is cast to extremes, of a certain kind. On one extreme end, action understood as a type of causation between molecules, events, and facts (and perhaps also properties, depending on how we understand them) assumes a notion of causation that is universally applied to all natural phenomena, whether they are agential occurrences or not. At the other extreme, causation of action is viewed as unique and radically different from all other causal types, where all of these non-agential types are potentially grouped together into a single overarching type: namely, that of event or fact causation.

2.1 Hanna and Maiese's 'essential embodiment theory': trying

In my first chapter, I reviewed standard causal theories of action and concluded that they cannot be unproblematically sustained given their numerous problems, including that of deviant causal chains. Hanna and Maiese suggest that the primary weakness of such theories is revealed in the fact that 'every classical causal theory inserts a vitiating metaphysical or temporal gap between antecedent mental causes and consequent body movements' (2009: 184). This 'gap' between mental causes and physical events presents two distinct problems, one of them unique to the concerns of action theorists, and the other holding much broader relevance in contemporary metaphysics and mind. The broader problem is that of mental causation, or, more specifically, the causal exclusion problem that continues to be discussed in philosophy of mind: given the subvenient physical base underlying mental causes (e.g., neurological processes), what room is left for the causal efficacy of the mental? The second problem pertains more specifically to action theory, and takes the form of the query: how is it that bodily movements, of the kind understood by us as implementing actions, can be definitively produced by antecedent mental causes without the worry of deviant causal chains being introduced?

These two questions are not necessarily connected, and one certainly need not entail the other. It is possible to entertain a theory of action that is susceptible to problems of deviant causation, without the difficulties of causal exclusion arising. Theories in which the problem of causal exclusion does arise, such as Hanna and Maiese's, may be formulated in ways designed specifically to avoid deviant causal chains, as we shall see. So the problems of causal exclusion and causal deviance can and should be treated separately. The question concerning causal exclusion does indeed receive treatment within a very vast literature that addresses it from a perspective entirely independent of questions concerning agency: for obvious reasons I will only be able to address it as a question pertaining to action theory. In this section I will evaluate a limited number of specific points contained in Hanna and Maiese's extremely comprehensive account, and offer some suggestions aligned with their general approach.

Hanna and Maiese argue that most causalist approaches, including Davidson's, are susceptible to both problems, causal exclusion as well as causal deviance. They present an alternative, non-standard causal theory of action that purports to avoid both of the above questions in turn. Their account merits interest for this chapter because they appeal to a contemporary variant of psychophysicality, known as 'embodied cognition' or 'embodiment theory' in the philosophy of mind, whose contemporary proponents have included Lakoff and Johnson (1999), and Varela, Thompson and Rosch (1991).¹⁴⁴ Embodiment theorists argue that the mental is never 'purely' mental: the mind, they say, is physically embodied, and is causally shaped by or dependent upon the body. Hanna and Maiese assimilate and develop this idea into their own essential embodied theory of action, which has as its primary aim that 'the essential embodiment of conscious, intentional minds rules out any *metaphysical gap* between the mental causes of action and the intentional body movements that are

¹⁴⁴ Hanna and Maiese claim that the tradition finds its roots in (among others) Kant, Merleau-Ponty, and later Wittgenstein, as well as Aristotle (2009: 15, 345-48).

their effects. This holds whether the metaphysical gap is generated by the substance dualism of agent-causal approaches¹⁴⁵, or by Jaegwon Kim's causal-explanatory exclusion problem for non-reductive materialist approaches that are based on supervenience' (2009: 197). In what follows, after explaining Hanna and Maiese's alternative causal theory, I will offer some remarks on both issues raised in this passage: (1) whether agent causal approaches are indeed all guilty of substance dualism and its consequential metaphysical gap, and (2) whether Hanna and Maiese's embodied theory of action escapes the causal exclusion problems faced by supervenience theories, or whether the similarities their own account bears to supervenience approaches expose them to the same charges.

Before continuing, I should make a clarificatory remark concerning the aim of this section given the background of the rest of this chapter. With regard to specific metaphysical solutions, this section leaves open the question of how the 'essential embodiment theory' should be explicated: the present section merely sets the stage for later introducing the concept of property fusion. Thus, the general approach of this section will be to clarify Hanna and Maiese's embodiment theory first at the 'macro' level of agents and actions, beginning with a discussion of their view on 'trying'. The succeeding section will then focus on the analysis of psychophysical properties ('property fusion') that constitute trying and action, on their view. The present section therefore complements the next: from describing the complex phenomenon of action and its causation, I will move to an analysis of the underlying constitutive aspects of that phenomenon.

Hanna and Maiese's non-standard causal theory of action begins with their acceptance of volitionism of a 'non-classical' sort, a view that combines elements from

¹⁴⁵ By the phrase 'substance dualism of agent-causal approaches', Hanna and Maiese certainly have the traditional Cartesian framework in mind, but as will become clear, they also equate the concept of agent causation with the Kantian notion of noumenal self.

O'Shaughnessy's earlier account involving 'tryings', and Frankfurt-style 'active guidance'. Classical volitionism is known for its pitfalls, several of which I will review shortly; but Hanna and Maiese claim to avoid falling into the classic errors with their non-standard approach. They identify as volitionists in light of their stated sympathies with O'Shaughnessy's notion of 'tryings' as constitutive of all action, successful or unsuccessful: they need not be 'effortful' or stressful, or 'full of striving' in the face of possible failure, as the term commonly connotes; for 'effortful' trying is defined as the kind of striving that occurs where there is a doubt of success. Trying is simply what agents do when they act, and a successful trying has satisfaction in a completed action, while an unsuccessful one does not: it simply remains an unsuccessful trying. In O'Shaughnessy's terms, 'Trying consists in doing, intentionally and with just that purpose, whatever one takes to be needed if, the rest of the world suitably cooperating, one is to perform the action' (1973: 369).

Whether 'effortful' or not, assuming these specially designated meanings, trying applies to all action: '[N]ecessarily *all* cases of intentional action involve effortless trying, even those cases that also involve effortful trying' (2009: 181).¹⁴⁶ Hanna and Maiese interpret 'trying' as synonymous with 'willing', the source or ground of all action (2009: 183). Classical volitionists likewise propose willing or volitions as the causal source of action, and arguably would not be opposed to accepting 'trying' (or some accepted definition of it) as wholly compatible with their approach. But Hanna and Maiese offer the further specification that tryings cannot be the kind of causal

¹⁴⁶ This statement no doubt sounds paradoxical. Surely the meanings given to these terms are unconventional and therefore possibly confusing, but a possible interpretation of Hanna and Maiese's statement may simply be that (i) wherever an agent does something, a concept of trying as intrinsically 'effortless' is also at work; (ii) even when an action is stressful or where there is little hope for success (i.e., 'effortful'), the fact that one does it *ipso facto* implicates the concept of trying as intrinsically 'effortless'. It must be admitted that such statements as the above make it seem as if in a difficult action (that requires more effort to accomplish) there are two distinct tryings at work, or perhaps two aspects of the agent's trying, i.e., one effortless and the other effortful.

source of action that has existence as a singular event occurring prior to the bodily movement it causes: they argue that tryings must be synchronous with the entire bodily movement, playing the role of active guidance, as Frankfurt proposed. The notion of ‘guidance’ according to Frankfurt is a commonsense one, approaching closely the intuitive idea that an agent possesses control over her actions (1978: 159). Frankfurt himself expresses as follows the motivation for rejecting a view of tryings as prior causes that end before the beginning of a physical movement:

A theory that is limited to describing causes prior to the occurrences of actions and of mere bodily movements...must inevitably leave open the possibility that a person, whatever his involvement in the events from which his action arises, loses all connection with the movements of his body at the moment when his action begins (1978: 158).¹⁴⁷

To avoid agential alienation of this kind, as well as the deviant causal chains that result, Hanna and Maiese distinguish themselves from classical volitionists in two crucial respects. First, where most volitionists assume that willing or trying is a mental act, Hanna and Maiese claim that ‘conscious, intentional minds are *essentially embodied*’ (2009: 184). Consequently, since trying is a minded movement of the will, tryings are also essentially embodied on their view.¹⁴⁸ Second, as already mentioned, they argue that ‘trying (whether effortless or effortful) and its active guidance (which

¹⁴⁷ Lavin makes a similar argument in relation to the notion of ‘basic action’ (2012), which I have already discussed in Chapter 1.

¹⁴⁸ On Hanna and Maiese’s view, effortless trying is ‘essentially embodied’, and it is the ‘unmoved motor of action’. They explain: ‘[E]ffortless trying is just the *same* as willing, and willing is nothing more and nothing less than an essentially embodied pre-reflectively conscious effective first-order desire that actively guides intentional body movements. So where intentional action is concerned, there is simply nothing *behind* our essentially embodied effortless trying and willing. Effortless trying or willing is just the unprecedented *ground, origin, or source of basic intentional acts*, a pre-reflectively conscious actively intervening mental cause that is also a synchronous active guide of covert neurobiological processes and overt intentional body movements alike. Effortless trying is what is ultimately *up to me*’ (2009: 183).

can also modulate in and out of merely maintaining guidance) is *synchronous* with the entire intentional action, including all the covert neurobiological body movements and overt body movements that necessarily and completely embody our conscious, intentional agency' (2009: 184).¹⁴⁹ Volitionists face the challenge that volitions, or willings, appear to be purely mental events that precede body movements, and thus seem to imply a metaphysical and temporal gap between mental events and the body movements. Lowe, for example, argues that volitions are 'individuated independently of the actions which are supposed to issue from them.... [I]t is one of the volitionist's central claims that volitions, or acts of will, can sometimes occur in the *absence* of the bodily effects to which they normally give rise (2000: 250). Volitionists are also forced to answer for a regress of acts that appears to open with the act of willing: if bodily movements are caused by a prior act of the will, then the act of the will must be caused by an act prior to *it*, and so on. Hanna and Maiese attempt to avoid these difficulties by arguing for a synchronous causal relationship between willing (or trying) and 'the entire intentional action'. For although 'trying', on their view, is similar to Lowe's concept of a volition in that trying, like a volition, can occur 'in the absence of the bodily effects to which they normally give rise', Hanna and Maiese reject that trying is a mental action.¹⁵⁰

For the same reasons that classical volitionism is rejected, Hanna and Maiese depart from O'Shaughnessy's conception of trying. As they characterize his view, O'Shaughnessy conforms to a classical volitionist approach in conceptualizing trying

¹⁴⁹ Hanna and Maiese explain overt and covert processes as having the following (common sense) differences: overt are external behavioral processes, while covert are internal neurophysiological processes (2009: 102).

¹⁵⁰ Although this point is not essential to the overall argument, it is not clear to me whether Hanna and Maiese's solution succeeds. They deny that trying is a mental action, and yet it cannot be conceived as something that simply happens to an agent: surely trying is something that the agent *does* (even when things end in failure). Perhaps they wish to say that even in cases of failure, trying is always embodied in some form of (failed) action, so that even in cases of failure, there is no distinct mental event of trying, but only failed action with which one's trying is synchronous.

as an ‘independent mental event¹⁵¹ that expires before the physical event of body movement begins’, whereas Hanna and Maiese define trying as ‘the irreducible mental *aspect* of an essentially embodied and therefore *essentially mental-and-physical* causal-dynamic living organismic process that therefore also has an irreducible physical aspect’ (2009: 185). In accordance with their revised definition, they identify as flawed O’Shaughnessy’s analysis of the ‘act-deception case’, in which a blindfolded man with a paralyzed arm has the mistaken belief that he has raised his (paralyzed) arm. Contra O’Shaughnessy, they write: ‘[T]he trying fails *not* because it is an independent mental event that fails to hook up causally in the right way with a later purely physical event, but instead because its essential embodiment in neurobiological movements has a causal-dynamic pattern that is in fact abnormally different from the one that would normally produce overt body movements’ (2009: 186).

O’Shaughnessy, it should be noted, does register his objection to ‘the claim that trying is doing what causes the event that gives its name to the action’ (1973: 369). There is reason to doubt that Hanna and Maiese’s characterization of him in their book is altogether accurate, for he later specifies: ‘Therefore, when [a] man succeeded in raising his arm, a psychological event, freely chosen and intended, *occurred side by side with* another event, a distinct event, a merely physical and act-neutral event, the event of arm rising’ (1973: 373, italics mine). Although his own clarification raises other problems (specifically, issues relating to causal exclusion, as well as what appears to be a ‘common factor’ assumption – see Chapter 1 for more on this problem), his remarks in this passage and elsewhere suggests that O’Shaughnessy’s concept of trying is *not* ‘an independent mental event that expires before the physical event of body movement begins’, as Hanna and Maiese claim (2009: 185).

¹⁵¹ The description ‘independent mental event’ is not made sufficiently precise by Hanna and Maiese. Presumably it means that the mental event (here, of trying) is conceptually distinct from the physical effects that follow.

O'Shaughnessy is indeed often criticized for having failed to secure a seamless metaphysical connection between tryings and bodily movements, however. He makes the infamous suggestion that because 'trying is *in essence* normally a cause of bodily change...[therefore] trying to move a limb is the mental "pineal gland"' (1973: 379).¹⁵² Hanna and Maiese's revised definition of tryings is arguably able to secure the proper 'gapless' causal relations by appealing to the notion of willing as an embodied event that is synchronous with 'the entire intentional movement'. Borrowing Frankfurt's idea of action as being actively guided all throughout by the agent's intention(s), Hanna and Maiese explicate the idea of embodied synchronicity in terms of actively guiding intentions. This revised definition of trying, they argue, avoids the problem of causal deviance, and thus overcomes any 'vitiating metaphysical or temporal gaps whatsoever between the mental activity of the agent and her intentional body movements' (2009: 103). If they are right about both embodiment of the mental and synchronicity with the neurobiological and behavioral processes constituting intentional movement, then Hanna and Maiese succeed in closing the metaphysical

¹⁵² O'Shaughnessy does however appear to recommend a psychophysical view in asserting trying 'as the pineal gland'. He writes: 'For although trying to raise an arm is *not*: whatever causes arm rising; it *is* and is essentially: an *x* which in the state of psychophysical normality, world permitting, is sufficient to cause arm rising.... Therefore trying, in being essentially a cause of a physical phenomenon *and* a linchpin of consciousness, serves a crucial bridge function between mind and body, not unlike that allotted by Descartes to the pineal gland. It is the key point at which the essential mutual interdependence of mind and body can be openly seen to be part of the scheme of things. It is like a psychic promontory that all but juts into the physical world' (1973: 378). Regardless, Hanna and Maiese are correct to say that O'Shaughnessy falls short in his attempt at establishing trying as a true psychophysical 'bridge function', for he consistently regards trying as mental and distinct from the physical object that it causes (1973: 378-379), with results that are frequently difficult to understand, e.g., 'Thus, trying to perform a physical act falls between sheer physical movement and ordinary mental activity' (1973: 379).

and temporal gaps that threaten all standard causal accounts of action. Once the gaps are closed, the problems of causal deviance disappear.¹⁵³ They explain:

We believe that our own non-classical causal theory, grounded in trying and its active guidance, smoothly explains [what goes wrong in cases of causal deviance]. The unfortunate climber never actually *tries* to loosen her hold, and the trembling robber never actually *tries* to signal to his accomplices. So any body movements resulting from their beliefs and desires – or indeed from any other causal antecedent – by a deviant causal chain will automatically be both uncaused by the agent and also unintentional, and thus not be a counterexample to our theory (2009: 155-156).

To summarize: whereas O’Shaughnessy’s analysis has trying as a mental occurrence ‘occurring side by side’ with the bodily movement as an ‘act-neutral physical event’ (1973: 375), Hanna and Maiese are at pains to clarify that trying (1) is not merely ‘occurring side by side’ with the movements of the body but rather embodied by them, where ‘embodied’ is a term of art that will be clarified only later on by their thesis of ‘property fusion’, and (2) is synchronous with the ‘entire intentional movement’ rather than an ‘act-neutral’ (i.e., merely physical) event.

2.2 Preliminary implications for philosophy of mind

I would like to address two substantial points of interest that arise in the foregoing discussion of Hanna and Maiese’s revised definition of trying. First, there is the purported contrast introduced between (a) their account of synchronicity with embodiment and (b) an alternative interpretation, O’Shaughnessy’s, of trying as also

¹⁵³ ‘Or more briefly and imagistically put, the conscious intentionality of the agent and her intentional body movements fit together as seamlessly as W.B. Yeats’s *dancer* and her *dance*: “O body swayed to music, O brightening glance, /How can we know the dancer from the dance?”’ (2009: 103).

synchronous but ‘occurring side by side’ with bodily movements. It may be thought that the latter causal relationship might intelligibly be explained by certain appeals to supervenience. Supervenience is a tempting explanation for the view that O’Shaughnessy defends, which says that when an agent raises her arm a psychological event ‘[occurs] side by side with another event, a distinct event, a merely physical and act-neutral event, the event of arm rising’ (1973: 373). Generally, the supervenience theorist might explain this kind of statement by asserting a relation of asymmetric variance. Depending on further metaphysical stipulations, the expressions of this relation are numerous: philosophers may say that the mental event ‘depends upon’, is ‘realised in’, or is ‘constituted by’ a corresponding physical event.¹⁵⁴ Thus, in the event that an agent successfully raises her arm, the supervenience theorist might say that trying ‘depends upon’ (or is ‘realised in’, or ‘constituted by’) the event of her arm rising. There is indeed a surfeit of available variations and modifications of supervenience from which to choose. Instead of attempting to sort through all of them, I want to say simply that supervenient relations possibly explicate the aforementioned metaphor of parallel (‘occurring side by side’) psychological and physical events in the following very general terms, where trying is said to ‘sufficiently cause’¹⁵⁵ a bodily movement (assuming agential success):

[Mental Event]_A sufficiently causes [Mental Event]_B,

¹⁵⁴ Any of these general formulations, it is noteworthy, are furthermore compatible with identity relations or substance dualistic ones, i.e., as between trying to raise one’s arm (where one is successful) and the event of one’s arm rising.

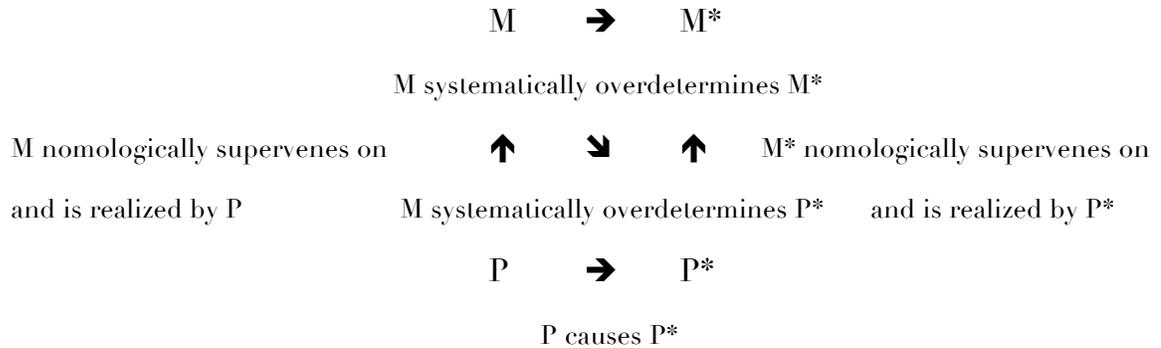
¹⁵⁵ The view that trying ‘sufficiently causes’ a bodily movement amounts to the neo-volitionist view embraced by both O’Shaughnessy and Hanna and Maiese. Given what I plan to say about event causation, there is much about this view that I find problematic. Since the purpose of this section, however, is simply to explicate some sense of ‘embodiment’ as preparation for discussing the concept of property fusion, I am assuming for now this terminology for the sake of argument. But I do address some of the problems of Hanna and Maiese’s neo-volitionism in the subsequent subsection.

where [Mental Event]_A supervenes on [Physical Event]_A and [Mental Event]_B supervenes on [Physical Event]_B.

The supervenience relations for both A and B hold, in accordance with the most general formulation of supervenience, if and only if a difference in [Mental Event]_A requires a corresponding difference in [Physical Event]_A, and likewise for [Mental Event]_B and [Physical Event]_B. For O'Shaughnessy, obviously, [Mental Event]_A represents the token event of trying that occurs when an agent wills to raise her arm, and [Physical Event]_A represents the token event of, say, neurotransmitters firing and neurological preparation that her body undergoes in anticipation of raising her arm. Were she to decide otherwise and alter her course of action, or desist from trying altogether, the set of neurophysiological events that form the subvenient base would be correspondingly altered. The bodily movement that results from her successful trying bears a similar analysis. An important implication of this latter view is that an agent's higher level decisions can alter the subvenient base consisting of neurophysiological occurrences, an implication that leads naturally to some concept of top-down or downward causation. Indeed, in last section of this chapter, I will examine a variant of this explanatory method in light of the assumption that downward causation is entirely compatible with supervenience.

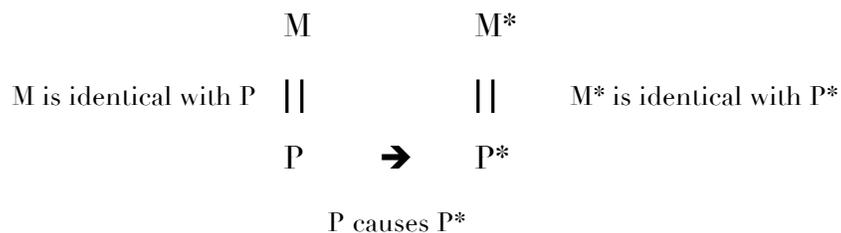
The events A and B could also be represented by mental and physical properties. For example, Hanna and Maiese propose a different possible account of supervenience along these lines, according to the following definitions (2009: 310): M = an event instantiating the fundamental mental property of my consciously willing to raise my right arm at time t_1 , P = an event instantiating the fundamental physical property of being the total state of my brain and body as I will to raise my right arm at t_1 , M* = an event instantiating the fundamental mental property of my consciously experiencing the raising of my right arm at time t_2 , and P* = an event instantiating the fundamental

physical property of being the total state of my brain and body as my right arm is raised at t_2 . They offer a (critical) diagram of the supervenience relations below:



Some explanation of the above diagram¹⁵⁶ is needed. The downward slanting arrow represents the causal relation between the mental property of an agent's willing to raise her arm and the physical property of her brain and body when the arm is then raised. Hanna and Maiese's representation of the supervenience relations is a critical one because they accuse the supervenience account of accepting 'causal overdeterminationism': both M^* and P^* are systematically overdetermined by M , in

¹⁵⁶ Their above picture is offered as a contrast from their illustration of reductive relations:



Hanna and Maiese remark: 'Thus Physicalism can indeed "save" the efficacious causal power of the mental, but *only* in the same sense in which, according to the notorious remark of a US general during the Vietnam War, sometimes the only way to save a village is to destroy it' (2009: 310-311).

their diagram, because the ‘fundamental physical supervenience-base P does all the causally efficacious work’ (2009: 310). This leads to the unwelcome conclusion that the mental event or property M (of the agent’s willing to raise her arm) ‘is merely along for the causal ride’, and is therefore ‘completely causally efficaciously otiose’ (2009: 310).

The explanation by appeal to supervenience has a very wide and complex literature that I am unable to undertake in full. I mention it here only in order to emphasize that Hanna and Maiese view their own approach as being self-consciously at a distance from the picture that supervenience theory presents, i.e., of a relation of dependence of mental on the physical, whatever more specific formulation one may adopt of these relations. There may be doubt that they succeed in this regard, however.

Supervenience is entirely compatible with synchronicity of the mental with the physical, so asserting such synchronicity does not *ipso facto* rule out supervenience. Perhaps it makes a difference that Hanna and Maiese insist that the event of trying is synchronous with the ‘entire intentional movement’ rather than with an ‘act-neutral’ physical event. Still, there could be a supervenience explanation of this picture, on which the mental correlates that are supervenient on the physiological occurrences that occur simultaneously to the agent’s intentional movement; the supervenient mental correlates could be understood as, perhaps, the (psychological) satisfaction of an intention fulfilled via the agent’s trying. So there seems to be no reason to reject supervenience as incompatible to Hanna and Maiese’s account of trying as the sufficient cause of the intentional movement in its entirety (i.e., as illustrated by Hanna and Maiese’s above diagram of the supervenience relations). The only possibility of Hanna and Maiese’s succeeding in their aim of distancing their account from supervenience theories, it appears, lies in differentiating their notion of trying as ‘essentially embodied’, and their proof that the relation of ‘embodiment’ is indeed a radically different approach from supervenience. I will consider their attempt in the following sub-section.

On the other hand, supervenience theories as applied to action are subject to a number of problems, the most challenging of which is the problem of causal exclusion. On the principle of causal closure, a physical effect must have a physical cause, which entails that the mental is excluded from playing an efficacious role in that effect. Supervenience is appealing for those desiring to avoid reduction of the mental to the physical, but its causal reliance or realizability in the physical basal level raises a similar but distinct question from that of causal exclusion. Even if the supervenient property is thought to be physical, its dependence on the subvenient base (understood variously as a relation of ‘realizability’ or ‘constitution in’ the physical base) makes it difficult to conceive of its independent ability to do any causal work at the level of the subvenient base. In other words, not only is the causal performance of the supervenient property potentially ‘excluded’ according to the principle of causal closure; the supervenient property is possibly also made redundant by definition of its dependent relations with the physical subvenient base. However the *mind* is conceived, the causal inefficacy of the mental, or epiphenomenalism, is traditionally regarded as untenable (Taylor 1927: 198; Davidson 1963: 9; Chalmers 1996; Hanna and Maiese 2009: 273). For most philosophers of mind, arguments entailing epiphenomenalism, or accepting compatibility with epiphenomenalism, are viewed as resulting in absurdity. The challenge of solving the problem of causal exclusion is therefore intensified by the need to avoid epiphenomenalistic conclusions. Jaegwon Kim’s arguments (1993) demonstrate convincingly that either identity theories (whether type-type or token-token) or strong supervenience of mental on physical properties leads to epiphenomenalism. His solution is, roughly, to embrace reduction (i.e., reductive materialism) on the basis that (1) only physical events can nomologically sufficiently cause physical events (causal closure of the physical), and (2) the fundamental physical properties of the natural world necessarily exclude inherent or intrinsic connections with fundamental mental properties. But supervenience

theorists would regard Kim's solution as, obviously, a defeat of their primary aim to preserve the conceptual distinctiveness of the mental from the physical.

Hanna and Maiese's response is to reject Kim's (2) but not (1),¹⁵⁷ in their metaphysical explication of trying as 'embodied', an explanation that they call 'property fusion' the topic of the succeeding section. For now, I simply want to reiterate that their revised conception of trying as embodied and synchronous with 'the entire intentional movement' presents a preview of the metaphysical challenges to come, as they need to convince us that their view really does distinguish itself from explanations appealing to supervenience, with all of its attendant challenges. The other point of interest promised earlier, in relation to Hanna and Maiese's revised conception of trying, is that their formulation specifically rejects 'act-neutral physical events', or 'mere' movements, when conceptualizing the effects or consequences of an agent's tryings. The effect of trying, when it is successful, and providing for ordinary agential and ambient conditions, is not a 'mere' bodily movement but rather 'the entire intentional movement'. This is a significant point of emphasis, I believe, because it opposes the causalist understanding that the event produced is a kind of 'act-neutral bodily movement' which, provided that the relevant psychological antecedents obtain, thereby counts as an action. Hanna and Maiese displace the causalist assumption by placing 'tryings' in relation to the whole, namely the action itself ('the entire intentional movement'), rather than in relation to another one of its parts, i.e., the 'act-neutral bodily movement' that could be (depending on the kind and presence of the psychological antecedents) voluntary or involuntary, an action or a mere movement.

¹⁵⁷ While they agree that only physical events can nomologically sufficiently cause physical events (causal closure), they argue that (i) the fundamental physical properties of the natural world do not necessarily exclude inherent or intrinsic connections with fundamental mental properties, and (iii) it is both metaphysically possible and also actually the case that fundamental physical properties include inherent or intrinsic connections with fundamental mental properties (2009: 297). Their argument here will be examined in the succeeding sub-section, on 'property fusion'.

The revised approach to tryings claims: wherever there is trying, there is (i) a completed action, granting normal agential and ambient conditions, or (ii) an incomplete action, e.g., cut short by the agent or by uncontrollable circumstances, or (iii) no overt action at all, e.g., in the case of paralysis. Where there is no trying by the agent, on the other hand, there may be (i*) involuntary, ‘mere’ bodily movement, or (ii*) no movement at all. In other words, in each case that the body of an agent moves, it is either an action (complete or incomplete) or a non-action type movement, i.e., something involuntary. There is no such concept as suggested by ‘act-neutral bodily movement’ that can be applied equally to analyse different agential movings, actions and non-actions alike. The concept of trying differentiates these movings as categorial types: for one type there is trying; for the other no trying is present. This latter implication of Hanna and Maiese’s account contains a distinctive flavor of disjunctivism, as reviewed in the first chapter, and thus it could be counted as an ally of that growing contemporary approach.¹⁵⁸

2.3 Difficulties

The latter two sub-sections have outlined Hanna and Maiese’s introduction to the topic of property fusion, via their concepts of trying and embodiment. I would like to conclude this section by enumerating a few tentative concerns for this introduction,

¹⁵⁸ Hanna and Maiese define their approach as a ‘direct realist disjunctivism in action theory’ (2009: 187), but as distinguishing different kinds of behavioral events across types of *tryings* (the successful cases versus the unsuccessful cases), rather than distinguishing between voluntary and involuntary (or nonvoluntary) movements. So they do not countenance (explicitly, at least) the version of disjunctivism for action that I have in mind here, viz. as distinguishing between the voluntary and the non- or involuntary movements. Their example contrasts a case of action with ‘act-deception’, e.g., wherein a blindfolded and paralyzed man mistakenly believes that he has raised his arm (through some external machination). They explain: ‘[A]lthough trying occurs in both intentional action and act-deception cases alike, just as sensory consciousness occurs both in sense perception and illusion, they are nevertheless categorically different types of trying, since their essential embodiment is categorically different’ (2009: 187).

which apply as well to the thesis of property fusion to follow. Throughout my explication of their account, I have assumed their conventional usage of the term ‘action’ as a species of ‘event’, as caused in the standard causal way by (a revised concept of) a psychological event or process, where the latter two terms, ‘event’ and ‘process’, are interchangeably used and thus assumed to be synonymous or conceptually equivalent. The first point of concern that arises from these assumptions has to do with the event causal account of agency that Hanna and Maiese require for their account. They assert that ‘the primary fact of causation is a real metaphysical relation between *singular* events in spacetime, such that a singular event e_1 causes a singular event e_2 which is not earlier than e_1 , under an intrinsic nomological constraint or law’ (2009: 259). In other words, their view admits of only event causation in the natural world. They write:

[I]t seems to us that in all and only possible cases of *natural causation in our actual world*, one singular event *lawfully necessitates* another event that is not earlier than the first event, in the following sense: Given the occurrence of the first event, it is *no mere coincidence* and *no mere accident* that the second event also occurs. The first event thereby *predictably produces* the second event. Therefore we adopt a [view]...according to which a cause is a nomologically sufficient condition of its effect. To emphasize this, we use the term *nomologically sufficient cause* (2009: 269).

Thus, on their view, all instances of natural causation are instances of event causation. The passage just quoted also reveals that they adopt the standard view that all causes are necessitating causes. Both of these assumptions can, and should, be critically examined: I will consider arguments opposing both in the third section of this chapter to follow (Steward 2012). A consequence of accepting them, for instance, is that Hanna and Maiese rule out the notion of agent cause as being anything other than the successive necessitation of events. Because they cling to an event causal account of

agency, the concept of agential trying can only be understood as a singular event that is synchronous with the event of the agent's intentionally moving.

I believe these assumptions merit further concern for two reasons. First, Hanna and Maiese are led to reject the idea of an agent cause on the basis of its constituting an 'unnatural' cause, i.e., one claiming exceptional causal relations among all other causal relations in the natural world, which are (they assume) event causal relations. They claim that the metaphysical gap produced by supervenience theories susceptible to the causal exclusion problem is likewise 'generated by the substance dualism of agent-causal approaches' (2009: 197), given that '[i]n agent causal theories, the alienation is the result of a vitiating *substance-dualist gap* into which a mysterious Cartesian *causal interaction* must be inserted between transcendent mental substances and fundamentally physical events' (2009: 105). So Hanna and Maiese simply assume that an agent cause must be a 'pure mental substance...which exists outside the series of natural events, and thereby is naturally *undetermined* by those events' (2009: 103), which brings about a 'fundamentally physical event'.¹⁵⁹

Hanna and Maiese identify a further problem with agent causal theories and the implication of a 'substance-dualist gap'. It is true that one should be wary of the appearance of any 'vitiating gap' between an agent and her movements, in which 'mysterious...causal interaction[s]' can be inserted, for such gaps are characteristic grounds in which deviant causal chains may arise (as argued in Chapter 1). Although Hanna and Maiese are right to argue that a minimal condition of philosophical adequacy for any causal theory of action is that it avoids deviant causal chains, should agent causal approaches simply be lumped together with all other causalist theories,

¹⁵⁹ In the background of this comment on agent cause, the influence of Kant's concept of noumenal self can be discerned. For example, they contrast their position from agent causal positions by writing: '[E]ffortless trying is the *unmoved motor* of action, or what is ultimately up to me, but *not* because it exists as a noumenal person-substance outside of time, as in classical agent causation' (2009: 183).

as they argue (2009: 105)? Is the attribution of Cartesian substance dualism to all agent causal approaches in general accurate? More caution is needed here, for not all agent causal theories need to presume a Cartesian interaction along the vitiating lines claimed by Hanna and Maiese. In the third section I will consider a variant of agent causation that does not presume exceptionalism about agent causes within the scheme of natural events, and does not conceive of such causation as being of a ‘purely mental’ sort. It is true that some notable agent causationists (notably O’Connor and Lowe) defend some form of substance dualism. But we need not accept the latter merely because we favor the idea of agent causes: merely positing an agent or substance cause does not entail that such a cause is ‘purely mental’, with effects that are ‘fundamentally physical’. For instance, the variant of agent causation that I want to consider in the third section proposes a conception of agent cause that might be described as both physical and mental. The reason it can be understood as both is because the agent is an animal: a living, sensing, and motile organism possessed of capacities that are simultaneously psychological and physical.

The second reason for raising concern over Hanna and Maiese’s account relates to their concept of trying as both ‘a singular event [that] lawfully necessitates another event’, and also an ‘essentially embodied’ event or process synchronous with the ‘entire intentional movement’ (§2.1 of this chapter). I want to register doubt that these two definitions are entirely compatible, as Hanna and Maiese seem to think. Classical volitionism, they argue, runs into difficulties because ‘it is an important mistake to think of trying as an independent intentional act that somehow has to be brought into being by another prior mental act of trying, on pain of infinite vicious regress’ (2009: 183).¹⁶⁰ In order to avoid the regress problem that follows this conception, Hanna and Maiese emphasize the idea of synchronous causes: the volition remains a singular event, but it may be the kind of cause that continues to exist for the entire beginning, duration, and end of its effect, e.g., the action. Thus they write:

¹⁶⁰ They attribute this view to Wittgenstein, *PI* §619.

On the contrary, on our account, effortless trying is just the *same* as willing, and willing is nothing more and nothing less than an essentially embodied pre-reflectively conscious effective first-order desire that actively guides intentional body movements. So where intentional action is concerned, there is simply nothing *behind* our essentially embodied effortless trying and willing. Effortless trying or willing is just the unprecedented *ground, origin, or source of basic intentional acts*, a pre-reflectively conscious actively intervening mental cause that is also a synchronous active guide of covert neurobiological processes and overt intentional body movements alike. Effortless trying is what is ultimately *up to me* (2009: 183).

There is much potential for confusion here. First, Hanna and Maiese assume that the volitionist's regress problem arises because of the idea that volitions must be temporally prior to and end before the commencement of an action. It is not clear that the volitionist must conceive of willing in this way, i.e., as temporally prior to physical movements. The idea that volitions cause willed movement does not necessarily lead to the distinct idea that volitions temporally precede willed movement: the latter requires the further assumption that causes are always temporally prior to their effects (or at least that, in the case of volitions, the temporal priority of causes applies for some reason).

Another problem with Hanna and Maiese's analysis is their presumption that the action must be caused in exactly the same way as the mental act is caused (and therefore the latter needs a cause, and so on). If this latter analysis is accurate, then the event causal story of agency arguably exacerbates the regress problem, for if one event causes another, it is justified to ask how the first event was caused. Alternatively, if we understood willing as playing a fundamentally different type of causal role from that of acting, the regress arguably would not develop. On this clarification, one would

still need to explain how volitions are not acts and yet also not wholly uncaused (i.e., in the sense that an agent is utterly passive with regard to her volitions), and how, if not wholly uncaused, volitions are also not movements that must be explained by further volitions. In addition to explaining how it is possible that willing is specially understood as an ‘unmoved mover’, one would also need to say whether the consequent understanding of bodily movement amounts to an exceptional cause among the natural causes.¹⁶¹ In any case, so long as the causation between willing and acting is understood to be of different kinds, it is possible for one to argue one’s way out of the threat of regress: the problem persists, *inter alia*, if causation is assumed to be homogeneous in kind. Insofar as Hanna and Maiese cling to an event-event causal story of agency, therefore, there is the possibility that the regress problem applies to their ‘neo-volitionist’ account of action as well.

Yet another potential area of difficulty resides in the lingering thought that Hanna and Maiese’s account appears remarkably similar to a supervenience explanation of Frankfurt’s ‘active guidance’ concept. One of the worries plaguing the supervenience account of Frankfurt’s suggestion is that it remains difficult to understand how a singular event synchronously causative of another singular event can be anything other than somehow external to the event effected, or perhaps as a merely simultaneous event that takes place alongside the action itself. Trying, on Hanna and Maiese’s account, is a form of active guidance of the body, which is to say that trying causes the body to move (under normal agential and ambient conditions). Trying is also, on their terms, a singular event that happens to occur synchronously with the intentional body movement that it causes. That trying is ‘singular’ apparently means, for them, that trying is conceptually distinct from the entire intentional movement,

¹⁶¹ I am not claiming that all of the classical volitionist’s ills would be cured with this move. There remain several significant problems if a volitionist (such as Lowe) desires to cling to a dualism of purely mental willings and fundamentally physical body movements. But this latter issue is, I believe, distinct from the threat of regress and its diagnosis.

but they wish to resist the conclusion that trying is therefore an external cause to the movement. The trouble here is that, despite their efforts to resist Frankfurt's picture, it is likewise difficult to understand how trying is a distinct event occurring synchronously with and causing the physical movement, and yet not something external to the effect (as causes are conventionally understood). Overcoming the image of a 'parallel lines' metaphor will require a great deal of metaphysical explanation, which Hanna and Maiese offer in a comprehensive treatment of their notion of 'property fusion', to which I turn next. Everything seems to hinge upon this concept, however, which for some philosophers will count as a strike against it, while for others it may provide a point of deep metaphysical interest.

3.1 The 'essential embodiment theory': property fusion

Hanna and Maiese's defense against the latter charge of similitude to supervenience will be answered, as I have said, in their metaphysical account of trying's 'embodiment'. My purpose in this section is to evaluate the success of their attempt of this latter aim. Hanna and Maiese's general argument is that 'trying' should be understood as more than a singular event synchronous to 'the entire intentional movement': trying is essentially embodied in neurobiological movements, with obvious psychological as well as less obvious physical aspects. Trying is an event that results in the neurobiological movements constituting action, but its causal relations do not follow the classical volitionist pattern of antecedent volitions and consequent movements, illustrated by the classic example of billiard balls striking one another. The event of trying, rather, is characterized as a synchronous cause of the intentional body movement; yet it is also to be understood as 'embodied', and therefore somehow dependent upon (or interdependent with) the very same physiological base that becomes the physiological process of action. How do we understand these rather abstract claims?

The metaphysical underpinning of Hanna and Maiese’s embodiment thesis is given as follows:

‘[T]rying is an irreducibly mental event whose fundamental properties are *fused* with the fundamental physical properties of a physical event...in the living organismic life of a certain animal, namely the intentional agent herself’ (2009: 185).

Hanna and Maiese name this thesis ‘property fusion’. Its principal philosophical foil is Kim’s claim, visited in the preceding section, that the fundamental physical properties of the natural world necessarily exclude inherent or intrinsic connections with fundamental mental properties (2009: 297). Kim’s arguments were made to bolster the materialist idea that any attempt to preserve the distinction of the mental from the physical, whether by availing oneself of identity theories or non-reductive supervenience explanations, would be bound to fail. Hanna and Maiese aim to directly refute his argument. They claim, ‘[I]t is both metaphysically possible and also actually the case that fundamental physical properties include inherent or intrinsic connections with fundamental mental properties’ (2009: 297). Property fusion, on their account, is a metaphysically possible and actual example of fundamental physical and fundamental mental properties being connected inherently or intrinsically.

There is further reason to believe, they argue, that materialism is epistemically unjustified: the history of classical and modern physics has a high record of failed theories, which serves by itself as strong grounds to think that even the current best physical theories are extremely likely to be subject to later revisions. They rationalize: ‘[S]ince we do not currently *know* what the nature of the physical world is, then we are in no position to assert with sufficient justification that the physical world is fundamentally physical’; and, quoting Chomsky: ‘[Materialism] will be a coherent position if its advocates tell us what counts as “physical” or “material”’ (2009: 302). If confidence in current accounts of materialism, as providing the only explanation

concerning the character of the physical world, is unjustified, then Kim's condemnation of mental and physical co-variance in general is not warranted either.

The historical failure of particular physical theories does not, *ipso facto*, provide sufficient reason for doubting that the world *is* fundamentally physical. On the other hand, Hanna and Maiese are right to question what is meant by 'physical': it is a concept that has been so dramatically altered since Aristotle's treatment of it in the *Physics*, that what is meant by contemporary materialists bears no resemblance at all to its ancient counterpart.¹⁶² I plan to say more about the modern concept, the potential conceptual divisions within it, and its relation to the other sciences, especially biology, in what follows (§3.4).

Whatever their motivation, Hanna and Maiese propose an intriguing alternative to both reductive materialism and supervenience. Their treatment of property fusion is comprehensive and complex. They begin by offering a seven-part description for their view:

Let us call a complete class of physical instantiation *I* of a given type of living organismic animal bodies *B* (say, human bodies or cat bodies) in a single type of compositional stuff *an embodiment E*. Then the refined version of mental-physical property fusion says this:

- (1) Under an embodiment *E*, an event or physical substance *X* has some fundamental mental properties M_1, M_2, M_3 , etc.

¹⁶² Aristotle's concept of the 'physical' was very broad, enough to include within its scope the particular substances that we would understand as animal agents, e.g., possessing powers of sense and action. Given the broadness of his view, there was no problem of consciousness nor mental causation that arose within his schema: psychological phenomena such as hearing or sight (or indeed action) were simply very natural occurrences associated with living, animal organisms. The latter is unsurprising given that for Aristotle, the psychological science is included under the title of 'physics'.

- (2) Under the same embodiment E , X also has some non-identical or distinct fundamental physical properties P_1, P_2, P_3 , etc.
- (3) For every M_i there is a one-to-one correlation with a corresponding P_i .¹⁶³
- (4) The members of each 1-1 correlated M_i - P_i pair are necessarily co-extensive.
- (5) The members of each 1-1 correlated M_i - P_i pair are not logically necessarily co-extensive.
- (6) The members of each 1-1 correlated M_i - P_i pair are mutually inherent or intrinsic structural properties of X .
- (7) X is a suitably complex living organism (2009: 354).

Let us first consider some general observations about these seven statements taken together. The last stipulation, (7), suggests that property fusion shifts the locus of the discussion as being slightly outside of contemporary philosophy of mind: as Hanna and Maiese describe it, their account more accurately represents a theory of living body ‘a suitably complex living organismic body’ rather than a theory of mind simpliciter (2009: 353).¹⁶⁴ Their account relies on a kind of functionalism that they call ‘living body functionalism’, which posits that ‘necessarily anything that has the same *naturally purposive self-organization* as a living animal body and also has all the same *causal powers* as a living animal body, is a living animal body. Or, in other words,

¹⁶³ From what I can gather from Hanna and Maiese’s complex account, M_i and P_i represent property types rather than tokens. Evidence to support this reading includes their commitment to a kind of multiple realizability, or what they call ‘multiple embodiability’: that ‘in principle the fundamental physical properties of the very same kind of living organismic animal body can *change* as it is instantiated in different possible compositional stuffs’ (2009: 354).

¹⁶⁴ They dismiss the consideration that there may be a metaphysical or dialectical disadvantage in calling property fusion a theory of body rather than of mind (2009: 353). This may be the case, but all the same their shift in focus will lead to difficulties in making themselves understood by philosophers of mind who cling to certain assumptions (for instance, specific notions of what counts as ‘physical’) in interpreting efforts like Hanna and Maiese’s.

necessarily anything that is the same kind of dynamic system as a living organismic animal body, is a living organismic animal body' (2009: 353). In sum, the focus of their account is on agents as living animal organisms, and something counts as a living animal organism if it is 'self-organized' (or simply 'organized') in a certain way, perhaps in virtue of its causal powers.

Another necessary, preliminary clarification concerns the phrases, in (1) and (2), 'fundamental mental properties' and 'fundamental physical properties'. Elsewhere, Hanna and Maiese define 'fundamental' in the following terms: 'a property *P* is a *fundamental physical property* of something *X* if and only if *P* is a necessary, internal¹⁶⁵ property of *X* and *P* is correctly attributed to *X* by at least one of the three basic natural sciences, viz., physics, chemistry, and biology' (2009: 23). The authors note that '[t]here is obviously some sort of explanatory circularity involved in defining *fundamental* physical properties in terms of the *basic* natural sciences. But in this context, the circularity seems benign' (2009: 23, fn. 8). Perhaps this circularity is benign, but explicating what makes a science 'basic' seems rather critical for understanding their definition of 'fundamental' properties, since 'basic' can be *prima facie* understood to be synonymous with 'fundamental'. In other words, the explanatory circularity involved in defining 'fundamental physical properties' in terms of what is 'fundamental' in natural phenomena needs to be resolved, even if such circularity is indeed benign.

¹⁶⁵ The concept of an 'internal property' is perhaps easier to comprehend; at least it is not circularly defined. On their view, a property *P* is an *internal property* of something *X*, if and only if the instantiation of *P* in *X* constitutes a *proper part* of *X*: 'For example, having a finger is an internal property of a human hand. But it is possible to have human hands that lack fingers. A property *P* is then an intrinsic property of something *X* if and only if *P* is a necessary, internal property of *X*. For example, having four sides of equal length is an intrinsic property of a square. Intuitively, it is an inherent property of a square that it has four equal sides. Understood this way, then, the terms "intrinsic property" and "inherent property" are synonyms' (2009: 23). On their account, intrinsic properties can be relational, as in the right-handedness of a hand, or non-relational, as in the whiteness of a piece of chalk.

Finally, an important but general issue that arises from this seven-part description of property fusion is the assumption that living animal agents are, in Hanna and Maiese's account, understood as existing on an ontological par with the events involving them. This is an issue that I will address later (§4); in particular, I will give reasons for why this concept of agents as ontologically interchangeable with events should be rejected, and consider arguments aiming to show that the causal power of the living animal qua agent is of an altogether different type from that of event causation.

3.2 Property fusion as a contemporary alternative: initial concerns

Hanna and Maiese's seven-part description of property fusion seeks to avoid both the identity relations defended by reductive theories and the supervenience relations of non-reductive approaches. On their account, the relationship between any given mental property, M_i , and its correlative physical property, P_i , is one of necessary co-extension but, as Hanna and Maiese emphasize, also non-identity. As for supervenience, the standard definition (assuming strong supervenience¹⁶⁶) says that supervenience relations hold if and only if a thing's P -properties are logically sufficient for the existence of its M -properties, and logically necessarily there cannot be a change in the thing's M -properties without a corresponding change in its P properties.¹⁶⁷ Although Hanna and Maiese's account of property fusion exhibits a kind

¹⁶⁶ I am unable to engage in a full discussion of strong versus weak supervenience here. Kim offers the following definition of the difference:

S[strong]: (Let A and B be families of properties closed under Boolean operations.) *A strongly supervenes* on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily* if any y has G, it has F.

S[weak]: (Let A and B be families of properties closed under Boolean operations.) *A weakly supervenes* on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A – that is, indiscernibility with respect to B entails indiscernibility with respect to A. (Kim 1993: 58).

¹⁶⁷ We may include under this heading Kim's definition of supervenient causation in terms of macro- and micro-events: "Macro-event m [e.g., a mental event] is a cause or

of co-variance and co-dependence between M_i and P_i (by definition), property fusion, according to Hanna and Maiese, ensures that two fused properties are ‘necessarily co-variant and necessarily co-dependent’ but also *mutually inherent or intrinsic structural properties of* whatever instantiates them, which is to say that they are mutually necessary relational spatiotemporal features of that sort of thing’ (2009: 305).

The success of property fusion’s being differentiated from either identity or supervenience relations relies on Hanna and Maiese’s argument that the non-identical co-variance, co-dependence, and co-extension between M_i and P_i amount to an M_i - P_i pair that are ‘mutually inherent or intrinsic structural properties of X ’. That is to say, as they argue, the seven-part description of M_i and P_i offered above exhibits a certain complementarity between them, i.e., one that explains their being ‘fused’ in a way that simultaneously avoids both the relations of identity and supervenience. Hanna and Maiese offer three examples to illustrate the nature of this unique relation. Their primary example is taken from applied geometry, and invokes the relationship between concavity and convexity, which are co-extensive, but not logically necessarily co-extensive. By the definition of qualitative or exact identity, if they were logically necessarily co-extensive, then concavity and convexity would be exactly (qualitatively) identical. But, Hanna and Maiese reply, ‘there is no *logical* impossibility in the thought that X is concave but not convex. It is certainly liberally conceivable...that there could be a world that consists entirely and intrinsically in a single infinite concave surface...and nothing else. So concavity and convexity are not logically necessarily co-extensive, and thus they are not identical properties’ (2009: 304). The other examples, on which Hanna and Maiese do not offer further comment besides merely noting them, are drawn from particle physics and biology: recognizing the ‘non-logically necessary reciprocal inherent or intrinsic structural relationship between the particle-

effect of event e in virtue of the fact that m is supervenient on some micro-event, n , which is a cause or effect of e ” (Kim 1989: 283).

position and particle-momentum in *quantum entanglement*', and the relationship between DNA structure and organismic structure in cellular life (2009: 305).

Let us focus on the case of concavity and convexity, which Hanna and Maiese argue represents an example of the complementarity found in property fusion. Their argument is that concavity and convexity exhibit a relationship of complementarity, not identity, with one another: precisely the kind of relationship between the mental and physical properties that constitute the event of trying or the 'entire intentional action' that follows. Supposing that Hanna and Maiese are right to assert that concavity and convexity are not logically necessarily co-extensive (and there may well be doubt about this assertion¹⁶⁸): are concavity and convexity, normally properties of a geometric line, 'complementary' in the same way that the mental and physical are alleged to complement one another in the event of trying? The authors do not explicitly address this question. Assuming their non-identity, concavity and convexity can be co-extensive properties of a single line if, it appears, that line is bounded simultaneously by two distinct spaces (perhaps an apt example is a well-known symbol from Chinese philosophy, the illustration of yin and yang, where the concave line that is the bound of yang is co-extensive with the convex line that is the bound of yin, and vice versa). What is striking about the fact that their co-extension depends on the same line being bounded by two distinct spaces, however, is that there is nothing specifically about convexity itself that simultaneously suggests its alleged

¹⁶⁸ Their claim depends substantially on the possibility of a 'hyperbolic or Lobachevskian world', which in turn relies on Hanna's commitment to APA logic, or 'the A Priori Argument Logic' (2009: 249). The details of APA logic are somewhat dimmed by their comment that '[p]recisely what sort of deviant logic the APA logic is, and the exact details of its formalization, do not matter for our purposes'; but a quick summation of APA is that it opens up 'the notion of logical possibility wide enough for it to let in liberal conceivability' (2009: 249-250). Given that their arguments concerning property fusion depend on a 'minimally acceptable non-classical logic' such as APA (i.e., upon which all of their a priori appeals to fused properties ultimately rest), this point in their argument could well be one where non-APA or classical logicians take issue.

‘complement’, i.e., concavity: the complementarity and co-extension is brought about by something external to the line, namely its two contiguous spaces. The mental properties underlying the event of trying, on the other hand, have been characterized as embodied, and therefore of themselves entail physical correlates, namely the physiological and nervous processes that an agent undergoes in preparation for acting.

A further question is left open by the assumption that concavity and convexity are properties (of a line) in a thoroughly robust sense of ‘property’: are they as robust, that is, as Hanna and Maiese likewise understand or hope mental and physical properties to be? By definition, the property of concavity is not identical with convexity, but especially where the two are co-extensive, there may be doubt that their existence as attributes of the same line serves more than an explanatory or descriptive purpose. The accuracy of the description, of the line as convex or concave, depends on the provision of a space contiguous to it: one could say that the description more precisely explains the bounds of that space (whether it is a ‘convex set’ or a ‘concave set’). However these geometrical implications are to be clarified, the following doubt remains. It may be that the mental attributes of performing an action (or of trying), in the intention or reason that the agent herself knows, and the correlative physical attributes exemplified in the neurophysiological events of the agent’s body at the moment of acting (or in trying), are real, metaphysical processes within the same agent; yet they are distinguished by significant differences that the concave and convex relationship lack. The difference between concavity and convexity, in contrast, might be thought as merely analytical or definitional.

3.3 Complementarity in action explanations

Setting aside these initial concerns, I want to give full consideration to Hanna and Maiese’s understanding of ‘complementarity’ in their account. They offer a further illustration in the case of action explanations, in response to the question ‘Why?’: that is, for a normal human being who has a reason to raise his arm and does so under

ordinary conditions, the question, ‘What caused your arm to go up?’ (or ‘Why did your arm go up?’), yields two possible answers:

(A₁): ‘I raised it.’

(A₂): ‘The neurobiological processes of my living organismic body [caused my arm to go up]’ (2009: 374).

It seems *prima facie* obvious that these statements are related to one another, as they pertain to the same event; but there is much disagreement over how they are related. The reductive materialist argues that they are identical insofar as A₁ simply reduces to A₂: A₁ simply *is* A₂ because the latter offers a more informative and complete understanding of A₁. The non-reductive materialist says that they are not identical to one another but that A₂ is metaphysically sufficient for the existence of A₁, provided that A₁ is ‘realized’ in or ‘determined’ by A₂: such is the case in which A₁ ‘supervenes’ on A₂ (however we understand the concept of supervenience). Hanna and Maiese wish to reject both of these explanations, arguing that both identity (reductive materialism) and supervenience (non-reductive materialism) entail epiphenomenalism of mental properties, or their causal inefficacy.¹⁶⁹ They argue instead that A₁ is related to A₂ as complementary properties are related to one another in the relation of fusion. Once more invoking the example of concavity and convexity, they write:

While either A₁ or A₂ is a perfectly correct answer to [the question], what is crucial from an explanatory point of view is that those two answers always go together, like the coupled vibrating strings of a two-string guitar. Given the essential embodiment of conscious, intentional minds, these are not *competing* causal explanations, precisely because they both belong to the *same* overarching complete and independent explanation. Similarly, one could cite the geometric properties of a certain material

¹⁶⁹ Or, for that matter, the elimination of mental properties, as some reductive theories argue.

curved surface considered as concave, or cite a different set of geometric properties of the same material surface considered as convex, but only in the context of a single overarching complete and independent geometric explanation that cites both of them together' (2009: 374).

Like O'Shaughnessy, Hanna and Maiese identify an arm's going up as having 'dual causal aspects', features which, as they explain it, entail that each aspect 'can be informatively cited in a causal explanation as *the* cause of the arm-rising, but only in the context of a single overarching complete and independent causal explanation that cites both of them in relation to one another' (2009: 374). Setting the geometric analogies aside for a moment, Hanna and Maiese's analysis of A₁ and A₂ as complementary explanations of action is compelling for several reasons. First, although it moves us away from the issue of properties (and the question whether or not properties have sufficient causal efficacy) and into the somewhat shallower realm of explanation, the explanatory complementarity offers the possibility that the statement, 'I raised my arm', is just as much an appropriate explanation of one's *action* in reply to the question 'Why?', as another explanation involving microphysical events. That the first explanation can be given equally in cases where the agent is entirely unconscious of 'the neurobiological processes of her living organismic body', suggests (without definitively confirming) that the two refer to distinguishable, but not wholly separable, processes. We could identify A₁ as the psychological process by which an agent knows her own intentions, or what she is up to in moving her body, and A₂ as the 'merely' physical processes of the body, the majority of which she has no knowledge except by scientific instruction.¹⁷⁰

The reductionist can reject this prospect, of course, by claiming that the physical processes of which the agent is unaware are nonetheless the very stuff to which A₁

¹⁷⁰ This distinction between these types of knowledge is famously suggested and developed by Anscombe (2000). There is much discussion about its proper exposition and evaluation (Haddock 2011, Setiya 2008, Thompson 2011, Moran 2004, Falvey 2000).

reduces. The reply, ‘I raised my arm’, is simply identical to (and explained by) the complex details involving the neurobiological processes of the agent’s body, whether she is conscious of those details or not. But this objection does not diminish the fact that the explanation of action persists alongside the explanation of neurobiological processes, and agents will overwhelmingly prefer (and often be exclusively capable of) giving A₁-type responses to the question ‘Why?’.¹⁷¹ That A₁ and A₂ are distinguishable, however, does not entail that they designate distinct processes operating wholly separately from one another. A₁ and A₂ ‘always go together’ in the sense that where one is correct, so is the other: this sounds like a kind of co-variance, but the relationship of one to the other is not of supervenient levels, but rather, simply, complementarity. This latter claim, at least, is what Hanna and Maiese wish to argue, and I will return to this point.

Second, Hanna and Maiese offer the interesting thought that the complementarity between A₁ and A₂ can only make sense in light of the fact that ‘both belong to the same overarching complete and independent explanation’. They never explicitly state what precisely this overarching explanation is; perhaps on their view it is simply the sum or aggregate of A₁ and A₂.¹⁷² I would like to extend their earlier thought (expressed by the final stipulation in their seven-part description) that property fusion is an attribute of a living body, or a ‘suitably complex living organism’, and suggest that the same overarching complete explanation to which A₁ and A₂ mutually belong is precisely the explanation that an agent is this kind of living, motile organism.

¹⁷¹ Obviously, responses of the A₁ type are precisely those sought when questions concerning moral responsibility, liability, praiseworthiness or blameworthiness, and other normative assessments, etc. arise. If philosophers wish to preserve as coherent these concepts, then explanations of the kind expressed in A₁ must be retained as being both explanatorily useful and arguably also metaphysically significant.

¹⁷² One could see it as the description that incorporates both A₁ and A₂ as mutually interdependent descriptions, and whose complementarity exhibits an event, i.e., of my arm going up, as of a special kind, namely, an action.

Human beings are naturally the best candidates¹⁷³ for fitting within such explanations, being capable of expressing the kind of action explanation in A₁ and comprehending the physiological processes described in A₂. But non-human animals are by no means excluded, for although they lack linguistic articulation they are no less capable of action bearing the mark of mentality (in varying degrees) as well as physicality.¹⁷⁴ Animal agents, both human and non-human, are mental and physical composites; there is no surprise that we ascribe both psychological and physical (psychophysical) attributes to them, and likewise, it is altogether understandable how explanations of the kind expressed by A₁ and A₂ should complement one another in agents. Wherever one is found as an expression by or characterization of an animal agent, so is the other.

This observation about the normalcy and pervasiveness of animal agents, as instantiating ‘embodied minds’, does not entail the truth of property fusion; but it does suggest that an account like property fusion, if it can be made more coherent, could very well explain the concept of an animal agent. Property fusion potentially

¹⁷³ Although they are not the only ones, as has been frequently adverted to in this thesis.

¹⁷⁴ As Anscombe succinctly puts it, ‘Intention appears to be something that we can express, but which brutes (which e.g. do not give orders) can *have*, though lacking any distinct expression of intention. For a cat’s movements in stalking a bird are hardly to be called an expression of intention. One might as well call the car’s stalling the *expression* of its being about to stop’ (2000: 5). My thesis sympathizes with her following of Aristotle (and Wittgenstein) in believing animals to possess intentions and therefore also the capacity for intentional movement (i.e., action), but I hesitate with regard to her denial that non-human animals altogether lack the ability to express their intentions. Obviously they lack it in the conventional human linguistic way, but it seems possible to speak of animal action bearing the mark or character of intention in a way that is often communicative. As examples, birds in the *corvidae* family or chimpanzees – and many other highly intelligent, sociable species such as orca whales or wolves – often teach their young skills specific to their ‘group culture’ (one might call it a social practice) by way of *demonstration*: arguably, a kind of action that has a communicative purpose, whose ‘intentionalness’ (to use Anscombe’s term) is obvious to the agent’s intended recipients.

explicates the mindedness of animal bodies by positing that the mental and physical properties constituting action are so closely correlated with one another (as expressed by Hanna and Maiese's seven-part description) as to be 'fused'. On their account, the fusion of these property types does not create a new fused property with novel causal powers; rather, the causal efficacy of mental and physical properties are individually preserved (thus ensuring causal efficacy of the mental, and avoiding epiphenomenalism).¹⁷⁵ Therefore, Hanna and Maiese's account of fusion has much less in common than might be supposed with Humphreys's earlier proposal, which suggested novel causal powers arising from a fusion relation between properties to support the idea of emergence (1996, 1997a, 1997b).¹⁷⁶ Their concept, it appears, instead remains close to the basic ideas of complementariness and co-extension: the 'fusion' of properties, on their account, describes the way the types of mental and physical properties involved in action and its trying 'always go together', just as the explanations A₁ and A₂ are co-variantly true. In both cases, 'fusion' (of the

¹⁷⁵ As already suggested, there remain difficulties with Hanna and Maiese's argument on this point. Not least of them is their analysis of property fusion as comparable to the relationship between concave and convex properties. It is not clear how far the analogy is meant to be taken: perhaps one accepts some notion of complementarity as playing a role in both cases, but in the case of action, the mental property of willing is understood as having causal efficacy in producing the physical property of one's intentional movement (in virtue of being 'fused' with the physical property of willing). It seems nonsensical to speak analogously of concave properties as causing convex properties, and vice versa.

¹⁷⁶ Another crucial difference of Humphreys's account is that the original fused properties, on his view, no longer exist (an Aristotelian might think of their relation to the whole as analogous with the relation of original elements within a homogeneous mixture). Thus he remarks: 'The key feature of [a fusion of properties] is that it is a unified whole in the sense that its causal effects cannot be correctly represented in terms of the separate causal effects of [the original property instances]. Moreover, within the fusion...the original property instances...no longer exist and they do not have all of their...causal properties available for use.... Some of them, so to speak, have been "used up" in forming the fused property instance. Hence, these...property instances no longer have an independent existence within the fusion. In the course of fusing they become [a novel] property instance' (1996: 61).

complementary kind, not Humphreys's) exists precisely because there is one 'overarching explanation', a whole living animal, of which the fused relata are attributes or parts.¹⁷⁷

3.4 Hierarchy and layer worlds

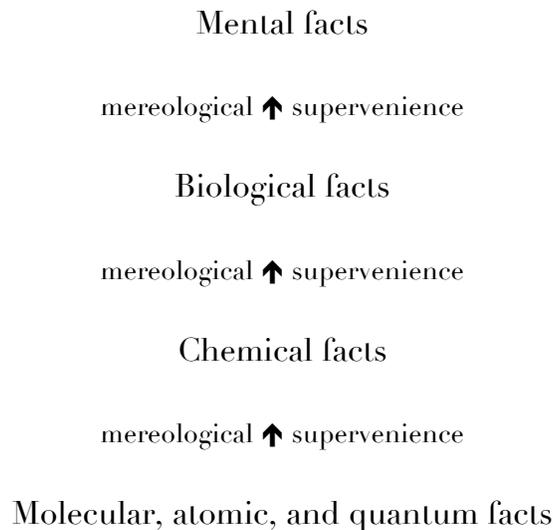
At the beginning of this exposition, I noted that a purported aim of this account was that it offered an alternative to supervenience theories. Whether Hanna and Maiese's account truly distinguishes itself from a general supervenience approach remains, at this juncture, undecided. One could argue that supervenience, with its relation of asymmetric variance and synchronicity (of the mental with its physical realizers) merely offers another way to articulate 'embodiment of the mind'. Furthermore, supervenience seems to have the added advantage (over property fusion) of assuming a hierarchical order where we normally suppose it to be relevant: in the idea that, especially where action is concerned, the fact of our mindedness (and animal mindedness in general) explains the motility of our bodies, or that the mental is understood as generally having governance over the physical.¹⁷⁸ That is, the fact that 'mind' sometimes governs 'the body' suggests a hierarchy that is not represented in the thesis of complementary mental and physical properties; but at the same time, one might argue, such a hierarchy does not necessarily force the property fusion theorist into giving up her basic commitment to complementarity between mind and body in action. Scientific studies suggesting the existence of synaptic malleability from

¹⁷⁷ Fusion, on Hanna and Maiese's account, represents the metaphysical articulation of their 'neo-Aristotelian hylomorphism', i.e., the contemporary idea that living bodies bear the character (or are constituted by) dual aspects, the mental and the physical. Whether the specification of these dual aspects understood now share great similarities with Aristotle's original thesis concerning form and matter as principles of, respectively, intelligibility and/or organization, and individuation, is an interesting question: however, it does not arise as a necessary matter of course in this chapter.

¹⁷⁸ Such top-down or downward causation is the topic of the subsequent section.

repetitive action,¹⁷⁹ and the development of musicians' or athletes' 'muscle memory' from habitual rehearsal, arguably bear out in concrete terms the hierarchical relationship of the mental over the physical. The proponent of property fusion, it seems to me, can still say that the evidence of neurophysiological interaction between mind and body in such habitual rehearsal is explicated by property fusion that corresponds to a supervenience-like hierarchy. I will say more about what this view might look like in the next section (§4). For now, there is no *prima facie* incompatibility, it appears, between the supervenience theorist's hierarchical premise and Hanna and Maiese's aim to establish metaphysical complementarity between body and mind in action.

Although they accept the relevance of such studies on synaptic malleability for their embodiment theory, Hanna and Maiese specifically reject the supervenience theorist's hierarchical premise, which they call 'the layered world picture' and illustrate as follows (2009: 316):



¹⁷⁹ For example, the brains of taxi drivers have been shown to have dramatic synaptic re-organization caused by the drivers' constantly learning new roads and routes. The brains of bus drivers on fixed routes, in contrast, show none of these synaptic changes.

They explain: '[T]he relation between the layers is one of asymmetric, non-reciprocal or one-way "upwards" modal dependence based on the part-whole relation: the higher levels are all ultimately either identical with or (logically or nomologically) strongly mereologically supervenient on the lower levels, in the sense that higher levels are entirely built out of smaller and smaller items occurring at the lower levels' (2009: 316). Their chief complaint with regards to this hierarchical structure is that it is vulnerable to 'downward decomposition of all entities and facts at any given level into mereological sums occurring at lower levels in the hierarchy' (2009: 317). This vulnerability is surely attributed to the assumption that the whole-part relation, of higher levels (with mental or psychological facts at the top) dependent on lower levels (with molecular, atomic, and quantum facts at the very bottom), is a 'one-way upwards' constitution relation. Or, as Hanna and Maiese put it, the relation is 'asymmetric' and 'non-reciprocal', which means, *inter alia*, that the top level facts are built entirely out of the bottom level facts: one could say (as many supervenience theorists do claim) that the top level facts 'emerge' out of the more basic or micro-physical facts. Asymmetry and non-reciprocity entail, of course, that the inverse is not true: according to the layered world picture, while the psychological and mental facts depend existentially upon the most basic physical level of molecular, atomic, and quantum description (with intermediary dependence on the natural sciences above foundational physics¹⁸⁰), the bottom level micro-physical facts do not in turn depend existentially upon the top level facts constituted by the micro-physical. Hanna and Maiese's main objection is that such one-way upwards constitution leads inevitably to downward decomposition, or reduction of top level facts to their most basic micro-physical constituents.

¹⁸⁰ Hanna and Maiese include in this illustration of the layered world account only two natural sciences, viz., biology and chemistry, that supervene on foundational physics. This list could be diversified. One distant possibility is that psychology might also be considered one of the natural sciences, or at least as fitting within the scope of natural and visible phenomena. I say more about this possibility in the next sections of the chapter.

Merely assuming a hierarchy of higher and lower levels does not, however, entail the threat of such downward decomposition. It is possible to understand the constitution relation between levels as not only ‘upwards’ but also ‘downwards’: such that the top level phenomena require basic micro-physical properties but also organizes the latter group’s existence as constituents of the supervening top level. The top level then can be understood as playing a causal role over its subvenient set; it exerts ‘downward causation’ over its own necessary constituents. It is not obvious that this kind of supervenient structure between the mental and the physical is incompatible with Hanna and Maiese’s account of property fusion. Their attempt to articulate the central notion of complementariness between mental and physical properties may have a correlate in the idea of downward causation of the mental on the physical, where the physical is also causally necessary for the mental phenomena involved (such as an intention-in-action). The kind of causal relation illustrated by this kind of view is not asymmetric but rather interdependent: both the physical and the mental are mutually necessary, as in property fusion, but the idea of mereological constitution is perhaps taken more seriously on this hierarchical account. We need not deny that the micro-physical parts are necessary for mental and psychological wholes, so long as there is a way to avoid the higher level entities being determined by their micro-physical parts, or the higher level phenomena being reduced to such parts. In the next section, I consider an account of agency that purports to do just that.

4.1 The causal exceptionalism of action

The preceding analysis of Hanna and Maiese’s account of agency highlighted, *inter alia*, their claim that ‘[i]n agent causal theories...alienation is the result of a vitiating *substance-dualist gap* into which a mysterious Cartesian *causal interaction* must be inserted between transcendent mental substances and fundamentally physical events’ (2009: 105). The result, they argue, is that agent causal theories present a mysterious or ‘unnatural’ causal relation where other natural causal phenomena appear

comparatively naturalized (and straightforwardly understood, usually in terms of event, property, or fact causation). This criticism of agent causal theories, that their proposed causal relation is metaphysically mysterious, is not entirely unfair.

Chisholm, for instance, describes agent causes as ‘unmoved movers’, a term that in medieval philosophy applies to divine being, and whose application to the theory of action evokes a notion of causation for agency that is unlike any other mode of causation in the natural world (1964: 32). Characterization of agency in such terms suggests that its causal structure is mysterious, perhaps even supernatural. Timothy O’Connor, another agent causationist, clarifies that we are ‘not wholly moved movers’, perhaps to express the thought that agent causation has unique standing among other causal types but does not yet approach the agency of God (2000: 67). The notion of an agent as cause of action is thus often justly charged with claiming that the causation of action is *exceptional* in the natural world: philosophers call this claim ‘causal exceptionalism’.

A common criticism of causal exceptionalism for agency frequently has its source in the assumption that all natural causation is *physical* causation of the same kind, i.e., of the kind that can be explicated fully in terms of fundamental physics (describable by molecular, atomic, and quantum facts). The idea that all natural causation is fundamentally physical causation presumes a ‘layered world’ view, as Hanna and Maiese call it (§2.2 of this chapter), in which a one-way upward mereological constitution of an agent’s mental and psychological aspects as supervening on the basic sciences (viz., biology, chemistry, and physics), ultimately results in mereological decomposition¹⁸¹ of such mental and psychological aspects to their fundamentally physical constituents. If it is true that the natural world and its causal phenomena are fundamentally physical (in the sense suggested by the very bottom level of the layered world picture), then there can be no agent causation in the inanimate world.

¹⁸¹ For now, I am assuming that such mereological decomposition is both an explanatory reduction as well as a metaphysical one.

According to such critics, the error committed by agent causationists is obvious: there can be no causal exceptions in the natural world, where the non-agential events of natural history – e.g., the pyroclastic flow surging down Mount Vesuvius, the trees shedding their leaves in autumn, the low atmospheric formation of cumulus clouds, etc. – exhibit a homogeneous type of causation. All natural phenomena can be understood, firstly, in terms of events, facts, or properties; perhaps these may be articulated in biological or chemical terms, but ultimately all can be explicated fully in terms of the fundamentally physical realm.

The question that Hanna and Maiese’s account raised in the preceding analysis of this chapter remains pertinent here: why should we assume that the natural world is fundamentally physical? At the root of this assumption is the inference that, since it appears that event causation most accurately describes most natural phenomena, one can conclude that event causation should prevail over *all* phenomena: but why should we believe that, especially in the face of potential counterexamples (e.g., ‘the sun melted the wax’, ‘the waves carried them away’)? If event causation alone necessarily explains all natural phenomena, then our assumptions about what counts as explanandum are likely to shift, i.e., to the level at which the phenomena to be explained are already in terms of fundamentally physical events. Put another way, we might ask: granting that we live in a material world, why should we assume that its ‘matter’ is ultimately all *physical* matter? After all, following the familiar division of the natural sciences, we can imagine that there is also a three-fold typology of ‘matter’:

- i. ‘Physical matter’, or P_p , which includes the following examples (from the micro- to the macroscopic): the behavior of elementary particles and waves, as well as virtual particles (the Casimir force, wave-particle duality, complementarity, radioactive decay), the interaction between high-energy particles and waves (superposition, wave interference, quantum entanglement), multi-atomic structures and optical fields (microscopic

interactions, temperature variable collision dynamics), properties of condensed matter (macroscopic interactions, superfluidity, superconductivity).

- ii. ‘Chemical matter’, or P_C , which includes the following examples (from the micro- to the macroscopic): elemental behavior (atomic charges, oxidation, chemical reactions and bonding types producing compounds), molecular properties (structure, acidity and basicity, phase states), substances (aggregate structure, energy, dynamic equilibrium between substances).
- iii. ‘Biological matter’, or P_B , which includes the following examples (from the micro- to the macroscopic): cellular phenomena (flagella movement, binary fission, mitosis, meiosis), sub-cellular processes (plasma membrane, nucleus, DNA, ribosomes, mitochondria), multi-cellular growth (flesh and bones, the nervous, digestive, hormonal, and respiratory systems of motile organisms, whole vertebrates and invertebrates).

It is a tautology that the natural sciences, however diverse in their scope of interest, are unified by a common interest in the material realm. The ‘fundamentality’ of the physical, as referred to earlier in the preceding analysis (§3.1), is alleged by reductionists to somehow encapsulate the whole of scientific knowledge, with distinctions between evidence types as holding no special significance. The above tripartite division of matter, however, presents a number of questions. First, why shouldn’t we think that there exists further differentiation between types of fundamentally physical properties of the type P_P , or the range of phenomena described by them? For instance, quantum entanglement describes a complex relationship (specifically, a ‘nonclassical correlation’) between separate quantum systems, seemingly of a vastly different causal relation from that between a downward quark and its proton. Second, why assume that the designation ‘fundamental’ belongs

to the class of ‘physical matter’ alone? Superposition is a feature of two wavelengths, for instance, and a fundamentally physical property of the property type P_P . Gaseousness is the fundamentally physical property of certain chemical elements, belonging to the property type P_C . The capacity for digestion is a characteristic of many multi-cellular, complex organisms, and a fundamentally physical property of the property type P_B . On this view of the ‘differentiated fundamental’, any talk of fundamental physical properties *simpliciter* leaves open the specification of ‘the physical’, whether qua biological matter, chemical matter, or physical matter. Why should we then assume that they are all ultimately of one material type?

A further question that arises from the above tripartite division concerns the diversification of matter: why stop at the class of ‘biological matter’? Why not diversify further to include the psychological sciences? The conventional scientific wisdom excludes psychology from the realm of the natural sciences, placing it instead among the so-called ‘human’ sciences, which include economics and sociology. But it remains unclear why it is that ‘psychology lacks the ontological authority of physics, chemistry, and the rest’, as Crane and Mellor have pointed out (1990: 187). Biology is fully accepted into the realm of the natural sciences, despite the dramatic and noteworthy contrasts between properties of the type P_P and properties of the type P_B , where particles and sub-particles serve as the subject matter for quantum and foundational physics, and empirically observable behavioral stuff or growth patterns which constitute the lifespans of living organisms serve as the subject matter for biological study. Biological processes also lead to organismic life, and, in particular, forms of minded organismic life (including rational and non-rational animals), which exhibit several striking features. Among them are consciousness, intentionality, and the capacity for intentional movement and its trying. These attributes fall under the purview of the psychological sciences, and they are of a kind to be dramatically contrasted with the property kinds examined by quantum and foundational physics. Such contrasts may indeed lead us to believe that there is no single, unified concept

of the ‘fundamentally physical’. So, if in spite of the latter doubt, we can accept biological science within the realm of matter, why not also accept psychology?

My purpose in bringing these issues concerning ‘matter’ to the foreground is simply to question the grounds on which the critics of causal exceptionalism for agency typically stand. If there is to be a critique of causal exceptionalism for agency, as however I agree there should be, then it will be better found elsewhere than in the claim that all natural causation is *physical* causation, i.e., of the kind that can be explicated fully in terms of fundamental physics, of the property type P_p . In the remainder of the chapter, I would like to consider the view that causal exceptionalism for agency is false, not because all causation is fundamentally physical causation of the type P_p , but in light of considerations concerning a particular variety of causal pluralism.

4.2 Agent causes and causal pluralism

The account of agency that I wish to consider in depth is offered by Helen Steward in *A Metaphysics for Freedom* (2012). Steward’s approach denies causal exceptionalism without likewise rejecting the notion of agent causes. Thus, although it is false that the causation of action is unique among the causal phenomena in the natural world, the agent causationist, according to Steward, is right with respect to one claim: that the causation of action is totally unlike the causation by events, facts, or properties. That agent causes are ontologically dissimilar from these other types of causes, however, does not *ipso facto* entail that agent causes are unique among them. This is, crucially, because the kind of causation exhibited by living, motile animals (including, of course, human beings) is also exhibited by certain – although not all – non-agential processes and things (i.e., substances) in the natural world. Thus the causal efficacy of

substances or things is a widespread phenomenon in the material realm, and animal agency is an important exemplification of it.¹⁸²

An important preliminary issue concerns the way ‘agent causation’ or ‘agent cause’ is understood on Steward’s view. In contrast with the standard agent causal view on which agents cause their actions, she explicitly insists that ‘actions should not be thought of as caused by agents’ (2012: 217). Instead, an action is regarded by her as ‘an exercise of [an agent’s] power to make the body (or particular parts of the body) move’ (2012: 32). Steward does not deny that agents can be causes, however. ‘Actions should not be regarded by the agent causationist as things *caused* by exercises of agent causation’, for this explanation begs the question of what exercises of agent causation are in the first place. Instead, actions themselves *are* exercises of agent causal powers (2012: 199).¹⁸³ Agents are causes, then, not of actions but of ‘the movements and changes in his or her own body in which acting normally consists’, as well as ‘all kinds

¹⁸² As will become clear from Steward’s definition of ‘agent causation’, the kind of causation involved is of ‘higher level substances’, and not simply all substances at all levels. It may be that substance causation of a sort is also exemplified at the level of the microphysical, with particles acting upon other particles or other microphysical substances. If this sort of causal interaction counts as substance causation, then it is of the very kind that, I believe Steward would argue, needs to be better understood as often behaving collectively as an identifiably ‘higher level’ thing or substance, event or phenomenon (such as wheels, whirlpools, volcanic eruptions, or walks in the park). As it is, my usage of the term ‘substance causation’ or ‘agent causation’ will presume this higher level ontology rather than the possible interpretation involving microphysical substances or things.

¹⁸³ In calling actions ‘themselves the *active* events in which we need to be interested’, Steward evokes a distinctive Aristotelian line with regard to the way ‘agent cause’ should be conceived. The Aristotelian understanding of action is as of an incomplete fulfilment of a potential for one’s body to be in a certain way or state of affairs (*Physics*, 257b 8-13). I examine this conception of agency in depth in Chapter 2. The point of raising Aristotle’s definition here is simply to provide a reminder that, for the Aristotelian, action can in no way be conceived as an effect produced by agents who do so, obviously, by acting. The potential regress is avoided on Aristotle’s account by ‘eliminating the gap’ between agent and action: as Hanna and Maiese quote Yeats, ‘How can we know the dancer from the dance?’ (2009: 103).

of other movements and changes in the world at large' (2012: 200).¹⁸⁴ In exercising their causal powers to make their bodies move, agents thereby act.

It may be thought that there is a contradiction in claiming, along with Steward, that what agents cause are not their actions but rather their bodily movements, while also attempting to defend a disjunctivist position on which some of the movements of an agent's body are actions (and others are merely movements).¹⁸⁵ One cannot cling to the view that agents do not cause their actions while simultaneously claiming that agents cause their own movements, some of which (according to my version of action disjunctivism) are actions. If one of these propositions must be jettisoned, then it would need to be the former: for it is unclear that such a view should pose a significant difference to the action disjunctivism under review, i.e., whether what the agent causes is her action or certain bodily movements, which may constitute an action. Perhaps, however, there is a way to ease the apparent contradiction. Although Steward denies that actions are identifiable with movements (2012: 33), she clearly regards as unproblematic for her account the claim that action 'normally consists' in 'the movements and changes in [an agent's] body' (2012: 199). There is no apparent contradiction in simultaneously asserting that agents do not cause actions and that the bodily movements that agents do cause normally constitute action qua an exercise of

¹⁸⁴ Hornsby, writing earlier, defends a very similar view. She argues that, 'in order to defend the causal role of agents, there is no need to say that they cause actions. Indeed there is every reason to say that agents do not cause actions.... What agents cause...are not the events that are their actions, but the effects or results in terms of which their actions may be described. And when we think of agents causing things, we don't think of them imposing themselves in causal chains that lead up to their actions' (2004: 17-18). Hornsby's rejection of such 'causal imposition' is part of her denial that agent causation necessarily represents an 'encroachment' among events, states, and processes, the usual causal candidates for understanding action. Her view, like Steward's, aligns more closely with an Aristotelian one on which agent and action are less separable than critics of agent causation typically believe, for Hornsby rejects the idea that agents and actions are 'fundamentally separate and distinct elements in the metaphysics' (2004: 17).

¹⁸⁵ Thanks to Adrian Haddock for raising this point.

the agential power to move: the important thing to maintain, as it appears on Steward's account, is that action does not *only* 'consist in' (or involve) bodily movements (so as to suggest an identity relation between action and movement), but that action is primarily an exercise of a causal power of the agent. If an action disjunctivist wishes to preserve the thought that agents do not cause their actions, I do not see why he or she could not propose similar stipulations on what is meant in saying that some bodily movements 'are' actions. The disjunctivist could clarify that bodily movements are likewise not identical with actions, but that *in* the fact that the agent is exercising the causal power to move her body, her bodily movements are therefore of a certain kind, i.e., the kind of movements that constitute action.¹⁸⁶ In any case, the issue here does not appear to pose a significant obstacle to the idea that the variant of action disjunctivism being considered may be integrated with Steward's approach to agent causation, to which I now turn.

Steward's argument for substance causation invokes the concept of 'top-down' or 'downward' causation, which may be understood as a way of organizing the higher level and lower level parts and processes that exist in whole living organisms or certain biological systems. With regards to agency in particular, top-down causation entails that 'the whole of an organized and integrated living system is able to affect the intuitively lower level processes that go on in its parts' (2012: 114).¹⁸⁷ Her approach to

¹⁸⁶ To be sure, thus far in the thesis this is not a clarification that I have seen the necessity to make, but I do not see how the disjunctivist view under consideration is blocked from making it: if, that is, one is compelled to insist that agents do not cause their actions.

¹⁸⁷ Steward's view therefore assumes, in one sense or another, the 'layered world picture' criticized by Hanna and Maiese, although she rejects the idea that the relation between layers must be one of one-way, asymmetric, upwards mereological constitution. This critical difference, I think, enables us to see her view as overcoming the threat of mereological decomposition that Hanna and Maiese raise in their depiction of the layered world assumption. In addition, Steward clarifies that her assumption considers the 'relevant levels in a *relative* rather than an absolute way; that is to say, that we need not commit ourselves to the idea that there are monolithic

substance causation distinguishes itself from other similar ‘top-down’ approaches. Many of these other approaches hang on the notion of ‘emergent properties’, which possess independent causal efficacy over the lower-level properties from which they are argued to emerge (Humphreys 1997, O’Connor 2000, Lowe 2000). Steward’s argument, however, is exclusively in terms of the causation of wholes versus their parts. Emergent properties are normally properties of a complex whole, but causation by emergent properties is not necessarily identical to the causation by the complex whole itself (2012: 225). Moreover, an important aspect of Steward’s argument, as will become clear, is her avoidance of the presumption that causation between properties is primary to causation by substances (inanimate things or animals), an aspect that further distances her approach from that of emergent properties.¹⁸⁸

Steward illustrates her case for substance causation by considering the phenomenon of a whirlpool. A whirlpool is the sort of system that involves the unfixed constitution of individual molecules of water, which are at one moment caught up in the forces of the whirlpool ‘system’ and at other times left out. Reductionists argue that the perdurance of the whirlpool can be understood entirely in terms of the lower level molecular arrangement, which changes from moment to moment. In other words, a ‘supervenience base’, or certain set of ‘basal conditions’, is entirely sufficient for explanations of apparently higher level phenomena such as whirlpools. But whirlpools

divides across the whole of nature of which there is a single correct account to be given, but merely that so far as the explanation of the workings of any given complex system is concerned, it may be useful to consider those workings as, in a certain sense *stratified*’ (2012: 227). Based on these comments and elsewhere, it is safe to say that Steward’s notion of top-down causation assumes a layered world view as an explanatory ‘metaphor’ rather than absolute ontological account, especially where animal agents are concerned.

¹⁸⁸ Although emergentism does not always entail exceptionalism, emergent theorists also frequently happen to be exceptionalists about emergent causation. They argue that human consciousness and free agency are uniquely emergent properties and *ipso facto* different in causal character from any other natural phenomena (Chalmers 2006, O’Connor 2000).

are generated by specific macro level forces, i.e., the flow of opposite running currents within a body of water, and these forces are what sustain the phenomenon of a whirlpool, regardless of which individual molecules of water are constituting the system from one moment to the next. Steward suggests, ‘to understand these forces and how they work, we do not look to each momentary individual supervenience base and consider how it generates the next. The persistence of the whirlpool is a phenomenon entirely blind to the details of individual molecules. It may indeed be a complete accident that any given individual molecule is part of the supervenience base of the whirlpool at any given moment’ (2012: 241).

Although the whirlpool is ‘a phenomenon entirely blind to the details’ of its lower level parts, much as my walking home is a process in which I am usually entirely unconscious of the details of the physiological, neurological and microphysical processes that constitute my walking, the ‘complete description’ of a whirlpool or a walk necessarily includes reference to the lower level constituents or basal conditions that exist from moment to moment. Steward raises a puzzle of coincidence with regard to the ‘assumption that each momentary individual supervenience base necessitates the next’ (2012: 241). Nothing in the basal conditions themselves, she argues, in the complex arrangement of molecules and other lower level occurrences, reveals how or why these conditions have aggregated just so, with striking coordination, as to produce a whirlpool. She asks: ‘What explains the coordination, the collocation? What we need to know is how it has come about that all these conditions have managed to obtain so fortuitously together’ (2012: 238). Referencing the temporally prior basal set that generates the supervenience base of the whirlpool – such as the molecular conditions that initially occurred when two opposing water currents were forced together at the formation of the whirlpool – does not solve the puzzle of how these prior molecular conditions are themselves produced together, of how they are coordinated in such a way by themselves alone. Steward writes, ‘No matter how far back we go along the chain, if we never raise our eyes from these lower level events,

we will lack what is requisite to explain why these conditions are (at any stage of the chain) produced *together*. And yet this is what is extraordinary; it is what needs to be explained' (2012: 238).

A key aspect to understanding Steward's puzzle of coincidence (of these lower level events), according to her argument, is that the availability of a *sufficient condition* (or set of conditions) for certain phenomena does not by itself entail *necessitating* microphysical conditions or prior circumstances. The issue of sufficiency has at least two relevant implications for the analysis of top-down causation. First, we must ask what the sufficient condition is a condition *for*. It may well be the case that there is a sufficient condition, e.g., a prior microphysical movement at t_1 , for a particle x to be in just the right place at the right time: that is, within the formation and flow of a whirlpool at t_2 . Even so, the existence of *this* sufficient condition does not by itself entail that there exists an entire set of sufficient conditions for vast microphysical collection of particles to be co-located altogether, as a whirlpool, at t_2 . Even if one were to identify a sufficient condition for each particle within the whirlpool at t_2 , such a set of conditions would fall short of explaining how it is that the movement of each particle obtains together. Their simultaneous co-location is a coincidence for which there is no sufficient condition or set of conditions, at least not as far as one can see from the perspective of lower level (microphysical) events alone.

The second implication related to the issue of sufficiency concerns the concept of a 'sufficient condition' itself. One may believe that 'sufficient condition' means conditions that are *necessitating*: such that, granting such conditions for the occurrence of an event or phenomenon, none other could have occurred. Here Steward cites Anscombe: 'For "sufficient condition" sounds like: "enough". And one certainly *can* ask: "May there not be *enough* to have made something happen – and yet it not have happened?"' (2012: 240). Steward argues in the affirmative:

We cannot just help ourselves to the assumption that everything that happens is inexorably necessitated by some prior state of the world. We *do* know, of course, that in another sense of the word ‘sufficient’, anything that actually happens must have had causally sufficient conditions, i.e. conditions that were ‘enough’ to allow for its occurrence, which is to say that nothing strictly requisite for the occurrence of anything that actually occurs could have been lacking. But that is different from there being conditions in place such that nothing else could then *possibly* have happened (2012: 240).

Weakening the assumption that ‘sufficient’ entails ‘necessitating’ is particularly important for Steward’s account of top-down causation. For if the coincidental microphysical circumstances of a whirlpool are to be understood entirely in terms of necessitating conditions, all in temporal succession generating the next set of microphysical circumstances, then there is literally no metaphysical space for the causal workings of higher level entities, such as a whirlpool or water currents. Steward comments, ‘Here one can indeed see no gap into which a phenomenon like top-down causation might be fitted’ (2012: 240).

Steward’s puzzle of coincidence is thus exacerbated by the assumption that ‘sufficient’ entails ‘necessitating’ condition.¹⁸⁹ Her solution to the puzzle is, first, to recognize that the presence of certain sufficient conditions for phenomena such as a whirlpool does not *ipso facto* eliminate the availability of other sufficient conditions – or, indeed, the possibility that no satisfactory answer to a particular event or phenomenon exists at all. Steward’s point is aptly illustrated in her reflection on Aristotle’s example of a

¹⁸⁹ The larger point that Steward wishes to make, that I (out of necessity) omit to mention in this chapter, is that the assumption of ‘sufficient’ as ‘necessitating’ is precisely the argument of determinism, against which Steward devotes most of her book’s efforts. I cannot do the account she offers justice within the scope of my thesis, although for my purposes I presume a concept of ‘sufficient condition’ that does not entail ‘necessitating’.

man who visits a well at the ‘wrong’ time, and is murdered by thieves (*Meta.* VI, 3). To explain what happened to the victim, we have at our disposal a variety of sufficient conditions for his being at the well at just the same time as his assailants. ‘But the mere existence of a sufficient condition does not by itself imply that the question how it was that these phenomena obtained *together* has a satisfactory answer or any answer at all,’ points out Steward. ‘However, in the case of the complex phenomena of which our world consists...we surely have the right to demand that there be such an answer’ (2012: 239).

The kind of answer she has in mind brings us to the second part of the solution to the puzzle of coincidence. The challenge that the case of the whirlpool poses specifically for the concept of action is, as Steward puts it, ‘to understand how on earth it can be that the animal [agent] has any real, independent efficacy *of its own*: any efficacy that does not merely reduce to the efficacy of its various parts’ (2012: 227). The argument that the animal agent *does* have independent causal efficacy, just as the whirlpool exists ‘entirely blind to the details of individual molecules’, purports to offer *the* answer to the question of how the microphysical phenomena all happen to obtain together in an apparent coincidence – wherever there are actions like one’s walking or complex non-agential occurrences like whirlpools. The provision of a higher level perspective counts not merely as an explanatory point in favor of (higher level) substance causation, but is part of the ‘relevant *metaphysics* of causation’ (2012: 239). Agents are causes of their own movements, operating ‘independently’ of the causal succession of microphysical parts that constitute higher level action.

This concept of agent cause treats causation by animal agents as having ‘independent’ efficacy; it is fair to ask precisely what such independence entails. It surely does not mean that agent causes, or top-down causation in general, do not depend on the causal powers of their constituent parts. Top-down causation admits of the constitutive necessity of the bottom-level, microphysical parts that constitute actions

such as walking and systems such as whirlpools. The foregoing objection to the idea of sufficient conditions as necessitating sought to establish that, although there exists such dependence on microphysical circumstances as sufficient conditions, the latter conditions do not by themselves determine (i.e., causally necessitate) or explain how it is that causal powers at the higher level are possible. The causal efficacy of animal agents is ‘independent’, therefore, not in the sense that it does not depend on the causal powers possessed by animal agents’ parts; such dependence is simply not also ‘upwardly’ determinative of the animal’s movements. In arguing for the ‘independent’ causal efficacy of the animal agent, a kind of top-down causal power, Steward thus avoids a traditional charge leveled against top-down causationists: that the higher level causal efficacy of agents, unmoored from the reality of its lower level supervenience base, results in a kind of causal power so ‘independent’ from the latter as to be something absolutely non-physical or ‘ethereal’. In contrast, Steward argues:

Nothing problematically ethereal is...involved; just as a whirlpool could have no effects were it not for the molecules of which it is composed having effects, so the way is clear for the straightforward admission that an animal could do nothing unless its parts simultaneously did things. The action of an animal is indeed constituted in this way by the actions of its parts. But as with the whirlpool, the key to seeing how the action of the animal is more than the sum of the actions of its parts is...likely to be the idea that higher-level processes can dominate and dictate the evolution and distribution of certain lower-level ones, so that, as far as the important causal metaphysics is concerned, the explanation of how a certain complex neural state of affairs has come to be depends upon higher-level processes and ontologies. And as with the whirlpool, we avoid the puzzle about how there could be anything more to be said than what is given by the low-level neural story (or indeed than by what is given by stories told at much lower levels than the neural) by simply

refusing the assumption that each momentary lower-level supervenience base necessitates the next (2012: 244).

The latter passage aptly summarizes Steward's main argument for top-down causation, while also defending the ontological independence – an independence of an entirely 'naturalistic' kind – of animal agents' causal powers. So far, my summary of her account has neglected to mention how the causal efficacy of agents prevails ontologically in the face of another kind of reduction (distinct from the reduction to the causal efficacy of an animals' lower level parts); namely, reduction in terms of neurological or microphysical events. For supposing the standpoint of one who resists the foregoing argument that the higher level causal powers of animal agents are irreducible to the efficacy of their lower level parts (presumably since the account of top-down causation may be found to lack compelling force): such an opponent may well resist Steward's main argument out of the idea that the phenomenon of higher level substance causation ultimately reduces (both metaphysically and explanatorily) to causation between events at a lower level, i.e., the physiological, neural, or microphysical.

The final explanatory detail for how the independent causal efficacy of animal agents is possible, particularly in the face of this latter threat of reduction, resides in Steward's account of causal pluralism. Some philosophers have denied that events can be causes at all, thereby asserting *only* the causal powers of agents or substances (Lowe 2008, Ayers 1968). Steward eschews this route, instead adopting the view that 'causation' is a category of diverse concepts, under which we may accept not only a causation by events, which Steward calls 'makers-happen', but also substances or collections of substances¹⁹⁰, which she calls 'movers'. Her causal pluralism also

¹⁹⁰ Steward's understanding of 'substance' is inclusive not only of animals and persons, but also 'stones and masses of air and water...as well as some of the smaller entities that go to make them up, like molecules and ions' (2012: 212). My concern lies more narrowly where such substances are animal agents. It may be that lower level

includes causation by facts, which she calls ‘matterers’. Each of these causal types are ontologically independent from, and thus also irreducible to, one another. This is not to say that the causal types may not be related to one another in important ways, but Steward insists that

we must not get the categories *mixed up* and assimilate them wrongly to one another in the service of a chimerical uniformity. ... In particular, it is absolutely essential to recognize that some types of cause (both the movers and the makers-happen) are proper spatiotemporal *particulars* while others (the matterers) relate to *general* factors (properties, features, aspects, etc.) and we only make a horrendous hash of our causal thinking if we fail to recognize that we are interested, when looking for causes, in *both* things that are particular *and* things that are general...[for] causal thinking involves a concern *both* with particularity *and* generality (2012: 211).

Steward’s distinction is an important one: namely, that between the entities falling under ‘generality’, and the particulars that actually *do* something, i.e., in the sense of enacting changes (or movements). The main force of this important distinction, I believe, is that it permits us to see how substance causation is not just another ontological category apart from events and facts, but indeed a causal category that takes a sort of ‘primacy’ over the other categories. Sketching the strongest case for this ‘primacy’ of substance causation, and in particular the agency of animals, will be the task of the next section. The explanation for this primacy of substance causation will, it is hoped, prevent the causation by substances from being reduced, both explanatorily and ontologically, to the causation by events involving that substance, or a causation between facts about the thing. It may be, of course, that Steward’s

substances, if one does accept them as such, require a different kind of argument (from that pertaining to animal agents) for defending their independent causal efficacy apart from the causation of events or facts.

argument for the irreducible causal efficacy of animal agents already settles this question. The aim of this section, after all, was to sketch persuasively her case for the claim that agents, as settlers of matters¹⁹¹ and initiators of their own movements, can be causes of a kind also found pervasively in other natural but non-agential phenomena. The argument for top-down causation purported to provide this explanation. Establishing the irreducibility of animal agency to causation by the animal's parts, it is hoped, may have a further, stronger case in the integration of an embodiment approach with Steward's account of top down causation.

4.3 Criticisms, and an integrative approach

In the remainder of this chapter, I want to address some of the central challenges faced by the account of substance causation laid out so far. After settling these difficulties, I will sketch a possible integrative account, where some version of Hanna and Maiese's embodiment theory is seen as entirely compatible with an agent causal approach exhibiting top-down causation, of the specific variety defended by Steward.

Many theories of top-down or downward causation presume emergentism, which posits that emergent properties exert special, or 'novel', causal influence over the basal properties from which they emerge. These emergentist interpretations of top-down causation are susceptible to a number of objections, most notably Kim's (1999, 2006). It may be thought that the variety of top-down causation defended by Steward faces these problems as well, so it is worth reviewing them. Kim's criticisms are mainly directed at the emergentist thesis that downward causation by higher-level properties is dependent, whether synchronically or diachronically, on basal physical

¹⁹¹ I have not devoted any attention to Steward's concept of agent as 'settlers of matters'. There will be further space in the subsequent section to address the issue, which offers the important thought that causation by agents involves a 'discretionary aspect' (2012: 246). Steward explains: 'It really is *up to the agent* which selections are made from the large repertoire of possible actions available: the agent counts as the settler of what is to happen with her body and thereby as the settler of what will happen to those parts of the world on which her body is able to impinge' (2012: 246).

conditions that serve as micro-constituents (again, whether synchronically or diachronically).¹⁹² The synchronic dependency is ‘viciously circular’, he argues, in that the emergentist claims that what produces higher level mentality, the basal set of physical properties, is simultaneously altered by the very mental properties that the basal set produces, and so on.

It may be thought that the variety of top-down causation defended by Steward escapes the brunt of these criticisms, primarily because her view is in terms of parts and wholes rather than emergent properties. But Kim’s objection – that the basal conditions (or the bottom level parts) are causally responsible for the whole that the parts constitute – may also apply to Steward’s version of top-down causation. Whether emergence is assumed or not, a similar problem seems to arise for the mereological version of downward causation: for the whole is causally dependent on the parts that constitute it as a whole, and simultaneously, claims the substance causationist, the parts somehow causally depend on the whole of which they are parts. Vicious circularity arguably results, as in the emergentist variation. For it seems absurd to think that the parts of a whole cause it to exist, while simultaneously the whole that depends on these parts also causes *them* to exist, as parts of the whole. Reduction is a tempting solution, in the face of this circularity. Should we not admit that a thing’s parts fundamentally determine the thing itself, and that this ‘upward determination’ is

¹⁹² Kim’s reply to the diachronic case, as opposed to the synchronic case, is too lengthy and tangential to be fully summarized here. As will shortly be seen, the synchronic case is accused of producing ‘viciously circular’ causation, given the simultaneity of downward and upward causation implicit in the emergentist’s account. Because the diachronic case explains the higher level causal exertion as occurring at a later time from the moment of upward determination, it escapes this criticism. However, since the diachronic case (like the synchronic version of emergentism) assumes the higher level properties to be mental properties and the basal conditions to be physical properties, Kim accuses diachronic emergentism of violating the principle of causal exclusion (2000: 318). Because Steward’s version of downward causation does not necessarily assume the causation by mental versus physical properties, the objection does not bear much weight in the present discussion.

more foundational than any form of causation one might otherwise introduce via ‘downward’ causation?’

The substance causationist may respond in a variety of ways to this objection. First, she could deny that the parts are constitutive of the whole in such a way that they causally determine the latter. A whirlpool, revisiting Steward’s example, is not upwardly determined by individual water molecules, since from one moment to the next the whirlpool is not constituted by the same collection of water molecules that happen to be taken up into the whirlpool’s current. That any particular arrangement of molecules happens to be a part of the supervenience base of the whirlpool could be, Steward points out, ‘a complete accident’, especially insofar as we grant that ‘we have no right to the assumption that each momentary individual supervenience base necessitates the next’ (2012: 241). Without this assumption, causal determination from below, or the upward causation of lower level parts on wholes, has no conceptual foothold: how is the determination of a constant (in this case, the whirlpool) possible when the assumed bottom level determinants themselves are absolutely variable?¹⁹³ So at least in the case of a whirlpool, the notion of ‘upward determination’ of parts on the whole can be set aside.

In the meantime, the question remains: does the rejection of lower level causal determination exclude *all* ‘upward’ causal influence? An extreme variation on Steward’s rejection of upward determination goes so far as to deny attributions of any causality relevant to the lower level parts exerted on the whole. The whirlpool surges on regardless of which and how many water molecules happen to be taken up into it as parts: its existence is wholly independent of the latter, entirely blind to them both explanatorily and ontologically. Perhaps one way to understand this approach is to think of a whirlpool as the kind of entity or system that simply happens to involve

¹⁹³ Although it is central to Steward’s inquiry, the topic of determination itself must be set aside as it is too large a topic to be dealt with here.

water molecules at each instantaneous time slice: no specific molecules of water are ‘necessary’ in the sense that the whirlpool would be conceived as causally dependent on those particular molecules of water. According to this extreme view, then, the whirlpool is a sort of thing whose parts exert no causal influence on it whatsoever, even as temporary constituents.

This view is unsatisfactory for the obvious reason that water molecules are not merely accidental features of whirlpools. There is at least one further problem with a view like this, however: it appears to downgrade the sense of parthood, in that it begins to sound incoherent to speak of parts and wholes at all. As mere accidental accompaniments to the whirlpool’s path, the water molecules that from one moment to the next happen to be taken up into the whirlpool itself apparently lack the status that parts or constitutive elements of a whole characteristically have. The obvious reply to this extreme position is that the molecular constituents of a whirlpool are important, *causally* important, for the higher level substance itself. Even when they are not strictly identified with the whole that persists through time – for the whirlpool is constantly exchanging its molecular parts for new ones, and artifacts such as wheels, although somewhat more stable, have parts that one by one may also gradually be replaced, as was the whole ship of Theseus – even then, the molecular constituents may be understood to have some kind of causal influence on the fact that the whole is what it is. For example, the fact that the molecular constituents of a whirlpool are *water* molecules, even though they undergo constant change, explains why the whirlpool ceases to exist when temperatures rise above water’s boiling point or falls below its freezing point. One could say that constituent parts, qua parts of a whole, are causes of another kind.

The more moderate approach to the objection of upward causal determination, however, still faces the earlier question posed by Kim, concerning synchronic circularity. The problem is that the substance causationist cannot maintain causal

priority of the whole, while simultaneously claiming that its parts exert upward causal influence on the whole. The only way to solve this problem, I think, is to take seriously the idea of causal pluralism invoked earlier in Steward's argument for her variety of substance causation. Her explanation is that substances, or 'movers', are causes in an entirely different way from facts, or 'matterers', and events, 'makers-happen'. That is, there are certain kinds of effects that simply cannot be attributed to facts, and other kinds of effects that cannot be attributed to substances.¹⁹⁴ In order to deflect Kim's criticisms, however, we must add at least another causal type to Steward's list.¹⁹⁵ Specifically, we should consider the category of 'constituent parts' or the 'stuff' that makes up a whole: whatever we call it, these do not cause the whole to exist in the same way that the whole causally organizes *them* specifically as parts.¹⁹⁶ Constituent parts are causes insofar as they make up a whole, but they make it up only insofar as the whole takes them in and organizes them as such.

Kim's criticism holds weight only if we presume that the causal priority of the whole over its parts is the very same kind of concept or causal relation as the upward causal exertion that constituents have on the wholes of which they are parts. Substance causation claims that the whole causally governs its parts – as a cat slinks towards a mouse, or as water funnels continuously off the coast of Suriname – in a way entirely distinct from the causal relation that constituent parts have with regard to their whole

¹⁹⁴ Steward deflects a potential charge that the difference between the causal types is merely explanatory; on her view, the differences between them are ontological as well as explanatory. She explains: 'Substances are simply causes of a different ontological type than either makers-happen or matterers, and it is entirely unsurprising that they do not play exactly the same role in explanation as causes of these other sorts' (2012: 220). For her, then, it is because of an ontological difference first that explanatory differences can be seen.

¹⁹⁵ One could arguably add many other different causal types to the list, although it is unnecessary to plumb the possibilities here. For instance, Anscombe points out that the removal of a doorstop from underneath the door it is jamming may be one such addition, a *causa removens prohibens* (Anscombe 2005, 'The Causation of Action').

¹⁹⁶ This idea may recognizably have its origin in Aristotle's material cause.

matter. There is no problem of vicious circularity for these cases because circularity assumes a directional conflict between identical ontological types.

The more moderate reading of substance causation does not, however, immediately escape the potential challenge that such causation is explanatorily redundant.

Steward's earlier argument concerning the puzzle of coincidence (§4.2) was intended to meet the challenge apropos constituent parts: i.e., that higher level substance causation is theoretically redundant as a proposed conception of 'cause', given that the lower level microphysical constituents and their efficacy constitute all that is needed to explain the movement of whirlpools and the agency of animals. Even if Steward's reply (§4.2) eases the trouble as regards the 'sufficiency' of constituent parts, the complaint of redundancy remains given the availability of explanations involving event causation. The popularity of Davidsonian type causalism in particular has ensured that higher level substance or agent causation is received with skepticism,¹⁹⁷ or viewed as explanatorily (and ontologically) superfluous. The Davidsonian reading, instead, encourages us to view events as the prime causal candidates. If it is true that my noticing a sign for a scenic detour causes me to impulsively take that detour, why must it also be said that *I* caused a significant alteration to the planned journey? Presuming that there is no issue concerning my responsibility for subsequently arriving late to my appointment, then citing agent causation seems like an added and unnecessary feature of what happened. Contrary to criticisms advanced by 'disappearing agency' theories (Velleman 1992), proponents of event causation need not be opposed to the idea of agency or agents, for there is no diminishment of the

¹⁹⁷ It has not helped that some proponents of agent causation have characterized it in somewhat mystical terms. Farrer, for instance, describes the lower level parts of an organism as 'the molecular constituents [that] are caught up and as it were *bewitched* by larger patterns of action, and cells in turn by the animal body' (1957: 57, italics mine). From Farrer's 1957 Gifford Lectures, published as *The Freedom of the Will* (1958).

sense that *I* took the detour path; it signals no conflict with my status as agent, in attributing the cause of my action to the event of my noticing the detour sign.

One way to counter the challenge of redundancy from the direction of event causalism is to deny that events are causes of actions at all (Lowe 2008, Ayers 1968). When it comes to actually getting something done, events simply do not seem to be the sort of things that do causal work. But, as Steward rightly points out, this is an extreme way to respond to the present challenge of redundancy (2012: 209-10). The event of my noticing a sign for a scenic detour does seem to figure causally in my subsequent action of adopting the detour route: the event of my noticing what I did explains why I went that route instead of another. If we assume the view that the ‘causal’ is broadly understood in terms of what explains or enlightens, an understanding that should have particular resonance for Aristotelians, then events as well can be causes of some kind.

Granting the latter need not amount to conceding that events are *the* causes of actions (even when it is specified, with sensitivity to the concepts of agents and agency, that such actions are *my* actions). Just as one potentially concedes too much in saying that the molecular constituents of an animal cause *it* to exist, one likewise grants too much status to events in allowing that they are the causes of actions. Events are better understood analogously: as the kind of causes that constituent parts and facts are. Facts are causes only insofar that they may play an explanatory role for actions. Constituent parts, in turn, are causes only insofar as they help to make up the whole that organizes them as parts. Constituent parts are therefore causes only in a limited sense, limited relative to the paradigm case of (higher level) substance causation. We can see the limitation of constituents especially in the case of living organisms. For attempts to explain the causation of action by way of parts or properties involving the brain and body fall below the level of the animal agent as a whole: they result in explications of action in terms of sub-agential, even fragmented elements. Explication

of action in terms of facts and events achieves a similarly limited result, from either direction: facts and events involving molecular, neurological, or merely physiological changes in the brain and body result in a similarly sub-agential picture of action, while facts and events involving the agent's wider circumstances may sometimes result in a viewpoint so far above and outside the agent that one fails to see how action involves an agent at all.

In light of these considerations, perhaps events should be treated as abstractions: certainly, they sometimes are indistinguishable from facts, as when 'the plant being wilted on Saturday morning' is put forth as an event.¹⁹⁸ Even more straightforward events, e.g., 'the sun coming out after the rain', 'the commencement of the new academic year', 'the child's turning five', etc., can be understood abstractly, recast as facts.¹⁹⁹ However they are conceived, the critical issue here is whether such entities are responsible for necessitating action.²⁰⁰ They are not. Indeed, an important feature of the agent causal view under consideration is that we should deny that actions are *ever* 'necessitated' by entities even other than events, such as agents themselves. Given the pervasive equivocation between 'cause', 'necessitate', and 'determine', considerations thus far have suggested that we cannot unproblematically claim that agents cause or necessitate their actions, even on an agent causationist view. The agent causal view reviewed in the preceding sections has, instead, advanced the idea that animal agents are causes of their bodily movements, not as necessitating or determining causes, but

¹⁹⁸ Steward characterizes this sort of example as following a liberal conception of events, arguing that we should cast doubt on whether they are true events or merely facts masquerading as events.

¹⁹⁹ A great deal could be said on the various linguistic ways to understand these statements. Steward argues that insofar as events behave as facts, neither are these the sorts of things that can cause actions (2012: 221).

²⁰⁰ Davidson's interpretation is that a relevant belief and desire pair (the event of an agent desiring an object *o* and believing relevant facts about *o*) are by themselves sufficient in causing physical movement, where 'causing' is in the sense of necessitating, and the physical movement being caused (provided that it is caused in the appropriate way) is also identified as an action.

simply as ‘settlers’ of matters.²⁰¹ As such, agents are causes par excellence. For the varieties of the concept of cause apropos action, and the limited influence that the sub-agential (or extra-agential) forms have in bringing it about, generate the demand for a higher level whole that explanatorily ‘organizes’ the different causal types of event, fact, constituent part, and property; its organization of these is such that they may count as playing causal roles (especially in the explanatory sense) in action, and without the ontological gaps from which causal deviance arises. This organizing higher level whole for action, of course, is the animal agent herself. The threat of reduction, raised in the previous section, is overcome once it is understood that the sub-agential (or extra-agential) causal features of agency could not be intelligible, qua such features, without higher level substance causation, specifically of the agential kind. Why then should we assume that event or fact causation, or indeed causation by properties or molecular and other microphysical elements, are the starting points for explicating action?

Rejection of the latter assumption is perhaps easier said than done. Steward’s arguments involving the example of the whirlpool made the case against reduction, specifically of the mereological kind: i.e., against the challenge that the whirlpool as constituted by water molecules can be understood entirely in molecular terms, so that the whirlpool itself is recognized as, explanatorily and ontologically, a superfluous entity. Steward’s response to the challenge, as reviewed before, was to argue that the organization of the lower level elements do not make sense without an ‘organizing’ cause from the higher level whole. She states: ‘It is quite true that the lower-level arrangements (once we have them) are “all we need”, but the crucial question is how the requisite causally proximal lower-level arrangements are to be provided for in the first place’ (2012: 236). As her statement suggests, the success of reductionist accounts depend upon the presumption that the entities such as the agent’s microphysical parts

²⁰¹ See especially Steward’s Chapter 2, ‘Up-to-Usness, Agency, and Determinism’, in *A Metaphysics for Freedom* (2012).

(as well as the events involving the agent's brain and body, related facts and properties, etc.) are entities or elements that are taken for granted: it is assumed that 'we already have them', and as such they are entirely adequate to explain agential phenomena without Steward's question, of how they are 'provided for in the first place', having to ever arise.

Steward's argument is that her question *does* arise: for what dictates that the microphysical elements are parts of the whirlpool, if not the higher level forces of the whirlpool itself? If the argument is to succeed, not just against the challenge of mereological reduction, but also over the other explanatory routes favored by causalists, then I believe it must be shown that an agential cause is needed first to generate the kind of causation involving these other sub-agential or extra-agential features. Just as mereological priority is attributed to a system such as a whirlpool, agent causation should garner explanatory first place among its related causal explanations. In the case of animal behavior (including human action), it is natural for us to countenance first of all explanations concerning the agent as cause, e.g., 'The cat pursued the mouse'. An analysis in terms of events possibly follows from the latter explanation, as e.g., 'The event of the cat's catching sight of the mouse caused its subsequent pursuit of it.' But analysis in terms of events or properties cannot possibly be explanatorily prior to explanations involving the agent as cause: what I mean by this is that we could not see the reductive analyses as true, without first acknowledging the agent's causal work. For the events in which 'the cat catches sight of the mouse' and 'its subsequent pursuit' both take for granted that the cat does something, i.e., that an agent causal explanation is already in place. According to this line of argument, the explanation of agents as causes is irreducible because the supposedly reductive analyses rely on *it* for their own intelligibility.²⁰²

²⁰² To some extent, a Davidsonian causalist might recognize that he takes agential explanation for granted. (Davidson thinks it is simply obvious that mental events

These latter remarks perhaps signal a somewhat stronger thesis with which Steward may or may not be in disagreement. I suspect, however, that her account tends toward this stronger thesis rather than simply positing that agential causes are simply one causal kind, irreducible among others. But does her argument really lead to this stronger conclusion, that agent causality is always explanatorily prior to other causal varieties, let alone establish this conclusion? There may be doubt that Steward's puzzle of coincidence proves what it purports to prove: that a higher level substance cause is necessary for the lower level microphysical elements to come together, in the first place, as an arrangement recognizable as a supervenience base. What does the puzzle of coincidence show, but at most that the arrangement of microphysical entities requires *some* explanation: *any* plausible explanation, which may or may not cite higher level substance causation? Perhaps her arguments against the main alternative, i.e., the necessitated succession of microphysical elements from one moment to the next, are compelling enough for one to assume that there could be no other explanation (for the phenomenon of coincidence) but that of top-down causation. But one may doubt that the latter represents a positive case for substance causation, let alone establishes it as the explanatorily primary situation from which all other varieties of causation are derived.

belong to agents; for this reason the criticism of disappearing agents does not impress him.) The problem for the causalist is that agential explanation does not reveal anything: it does not offer any kind of explanans, but is rather the explanandum, the thing to be explained. So the Davidsonian causalist needs to be persuaded that the concept of an agent as cause (i.e., that explanation in terms of agent cause) is intelligible and not merely a trivial re-statement of the problem. Some proponents of agent causation (e.g., Lowe) will face a very different problem: they will accept the explanation of agent causation as intelligible and informative, but reject the possibility that explanation in terms of causation by events or properties can be made available without damaging the view of agent as cause. It seems to me that a successful account of agent causation should simultaneously address both of these objections, the Davidsonian causalist's and the agent causationist.

That the account of top-down causation given by Steward is thus limited, I believe, may possibly be diagnosed as an implication of her original characterization of animal agents themselves. The central notion defining what it is to be an animal (agent) is that of ‘settling’: as Steward explains, ‘it is natural to think of such animals [as cows and sheep] as the *settlers* of various matters that concern the movement through time and space of their own bodies’ (2012: 75). The concept of an agent, on her account, has the following additional features in addition to the idea of ‘settler of matters’ (2012: 71-72):

- (i) an agent can move the whole, or at least some parts, of something we are inclined to think of as *its* body;
- (ii) an agent is a centre of some form of subjectivity;
- (iii) an agent is something to which at least some rudimentary types of intentional state (e.g. trying, wanting, perceiving) may be properly attributed;
- (iv) an agent is a settler of matters concerning certain of the movements of its own body in roughly the sense [that]...the actions by means of which those movements are effected cannot be regarded merely as the inevitable consequences of what has gone before.

Although these four features do not wholly amount to a functionalist account of what it is to be an agent, the emphasis on agents as settlers of matters appears to be an overwhelming emphasis on what agents *do*. On one hand, this seems entirely appropriate given that the prevailing interest of Steward’s thesis lies in the concept of agential doings. From another perspective, however, we may be curious to know what kind of thing it is that possesses the capacity to engage in such doings. Obviously features (ii) and (iii) address this latter question to some extent. Furthermore, it may already be clear that features (ii) and (iii) in one way or another entail (i), if not (iv),²⁰³

²⁰³ See Brodie’s ‘Agency and Determinism in A Metaphysics for Freedom’ for a more precise formulation of this entailment (2013: 572). Brodie argues against feature (iv)

in which case we at least have some line of thought about what sort of thing an agent *is* for it to possess the capacity of moving itself, or settling matters concerning certain movements of its own body. Features (ii) and (iii) suggest that agents possess intentional or mental attributes, specifically of a kind normally associated with the ability to move their own bodies. This is not a line of thought that Steward lays out in special detail. Nor does it seem necessary for her to do so, however, so far as her overall aims are concerned: they include, *inter alia*, the conclusion that agents are truly causes, irreducible to the causation of their parts.

If, however, one wishes to establish the stronger thesis that agents are *explanatorily primary* causes, on which all other causal kinds depend (as far as the causation of action is concerned), then a similarly stronger argument is needed to link the functionalist definition of agents in features (i) and (iv) with the preliminary thought expressed in (ii) and (iii): i.e., agency exhibits certain mental aspects, specifically certain kinds of intentionality, as well as being the bodily movements of which agents are by definition capable. My proposal is to consider an integrative account, i.e. one that accepts Hanna and Maiese's embodied approach as entirely compatible with Steward's agent causal, top-down argument. As mentioned previously, Hanna and Maiese argue against such compatibility between agent causal theories and their embodied approach to action: their argument assumes, wrongly, that agent causal approaches (1) always entail causal exceptionalism, and (2) are uniformly characterized by an irreconcilable mental and physical gap. Steward's agent causal account explicitly rejects the first assumption; as for the second, there is no reason why the top-down causal analysis should sustain any mental and physical gaps. Her analysis merely shows that the whole animal, in acting, plays an organizational and thus causal role on its lower level parts.

being similarly entailed by the first three features, whether singly or together. This detail, while being significant for Steward's discussion on determinism, does not substantially alter my current point.

In §3.3 of this chapter, I raised a question concerning Hanna and Maiese's conception of the mental and the physical as being instantiated in a relationship of complementary properties: how do we make sense of these properties as co-extensive pairs that 'always go together', and yet also as non-identical, distinct concepts? Hanna and Maiese seem to suggest that the mental and physical are complementary and co-relational, just as agential explanations of an arm movement ('I raised my arm') and purely physical or microphysical explanations of that movement ('the neurobiological processes of one's living organismic body') correspond to one another under normal circumstances. Their explanation of the correspondence is that the two descriptions belong to 'the same overarching explanation': they do not clarify what precisely this overarching explanation is an explanation *of*, but presumably, it concerns at least the whole action. The argument for property fusion, on Hanna and Maiese's account, involves the attempt to adopt an identical or very similar relational structure for mental and physical properties and their natural co-relation within animal agents. Although the argument for it may seem somewhat opaque at times, property fusion holds the promise of explaining the concept of an animal agent, as minded, motile (and obviously material) beings, whose mindedness is characterized (by way of an *a priori* argument) as being embodied.

Some critics of Hanna and Maiese's account will undoubtedly fail to appreciate this *a priori* narrative. To them, there is perhaps nothing more to suggest than the obvious fact that animals are as pervasive in our universe as are their agential exemplifications, and that specifically what agency exemplifies is, *inter alia*, mentality (in its various intentional forms, including sensing, desiring, perceiving) as well as physicality. The potentially more difficult question is why mentality and physicality should be found together as they are, and no less in a 'complementary' relationship, as it is claimed. Here we have a kind of 'puzzle of coincidence', similar to the previous puzzle raised by Steward's example of the whirlpool, whose microphysical aggregates coming together just so to form a whirlpool posed a question for reductionists. I believe that Steward's

argument for top-down causation can be effectively deployed to help diminish some of the mystery that may surround Hanna and Maiese's conception of property fusion.²⁰⁴ In particular, where Hanna and Maiese's explication of complementarity depended on a rather problematic example of concavity and convexity (these difficulties were outlined in §3.2 of this chapter), Steward's account usefully advances the argument for animal agents by invoking the concept of causation by a whole. The relation between the causal whole and the causal disposition of its parts is made relatively clear by Steward's argument for top-down causation, whereas 'property fusion' falters in attempting to explicate property complementarity as the basis of 'embodied' wholes.

Top-down causation, as it appears in animal agency, *is* causation by the animal: to move its own body, to 'settle matters' concerning certain of the latter's movements, and to effect sometimes long-ranging changes in the physical world beyond its immediate self. One of the conclusions of Chapter 2 was that, in doing these things, the animal agent enacts a change: specifically, the animal actualizes a potentiality for being in some end state, the achievement of some state of affairs. The matter of which the animal agent is said to settle is clearly physical: it certainly involves the movements of the animal's body as it reaches for some food, and it also involves the object in its environment that has now become food consumed by the animal in its movements. There can be no doubt that the fulfillment of the animal's potentiality for reaching toward and consuming food carries certain psychological requirements as well: the animal perceives an environmental object as food, desires it, and reaches toward it (or

²⁰⁴ One sense in which Steward's argument may be used to assist Hanna and Maiese's account is in the latter's discussion of complementary explanations of actions as belonging to an 'overarching explanation' (§3.3 of this chapter). Hanna and Maiese do not explicitly say what this overarching explanation is that unifies the complementary explanations of action and simultaneous neurophysiological processes. Steward's account offers the possibility that the latter question may be answered with reference to the existence of the whole animal agent. I do not spell out this particular point in detail, but it seems to me a promising way to ameliorate some of the unsatisfying conclusions from Hanna and Maiese's discussion.

tries to) all at once. Even Davidsonian causalists agree with this general story (aside from their claim that non-human animals are excluded from it); the issue that distinguishes their approach from an integrative top-down causal, psychophysicalist account, is raised by the question (among others), of *what* causes the physical movement or behavior of the animal agent so described. The standard causal account requires only *that* the agent desires and believes in a relevant object of pursuit; only the agent's desiring and believing cause her body to move towards that object, a movement that *is* her action.

The integrative account does not deny attributing some causal role to the events of believing and desiring, while also broadening their potential application to the inclusion of non-human animals. The integrative account instead locates the error of Davidsonian causalism in at least two places: first, in the latter's identification of 'action' as the physical effect of necessary and prior mental events; and second, in its narrow conception of the causal antecedents of action. Hanna and Maiese recommend that the causal antecedent of bodily movement be identified in agential trying as embodied. Perhaps this thought was not made perfectly intelligible throughout the consideration of their account, but it is no less an intriguing thought apropos a substance causationist account of animal agency. Trying encompasses the events of believing and the desiring that characterize agency, but trying is constituted by more than these mental events. It is something that the agent *does* with her body, so it has both physical as well as mental properties. And, on Hanna and Maiese's account, *it* is what proximally causes a complete or incomplete action, presuming that the appropriate conditions for action are in place. The action is not identified as an effect of the causal antecedent, i.e., the agent's trying; rather, trying partially constitutes the action when the latter is successful. On this view, all action begins with trying, even if all tryings are not always fulfilled by the successful completion of an action. Trying is therefore essential to agency: it is the essential thing that animals do, with their capacities for sensing, desiring, and perceiving.

If the foregoing is acceptable, then it comes as no surprise that mentality and physicality are found together as they are – whether according to the property fusion theorist in particular, or in the view of a mere observer of the natural world at large. Top-down causation makes sense of the claim that there are mutually intrinsic or inherent mental and physical properties in trying and successful acting, because top-down causation (as it applies to agency) *is* causation by the animal agent in doing what it does. Most essentially, such causation is manifested in the animal’s tryings, in the sense defined throughout §2 of this chapter, to move its own body and thus enact certain changes in its environment, for an end state that it views as desirable. If we accept the property fusion account of mental and physical complementarity, we can accept it on the basis that top-down causation is a type of organizational cause. It organizes the parts of the animal so that, under normal conditions, they ‘behave’ in unified accord with the animal’s higher level aims. This unity of parts expressed in an animal’s tryings and actions reveals, *inter alia*, that the relevant mental and physical properties ‘always go together’. The complementarity between them exists only in virtue of the whole, living animal, whose primary mode of enacting change, via top-down causation, neither promotes nor prohibits such complementarity.

That such mental and physical complementarity does exist, however, counts as a mark in favor of the stronger thesis about substance causation apropos agency: that it is not merely another cause among many, but indeed *the* primary causal explanation on which the other causal types rely. Property fusion, or psychophysicality in general, is able to explain what kind of substance an animal is, an explanation that is unavailable from the perspective of the sub-agential causes. Understanding the concept of an animal as the unity of psychological and physical processes or changes, on the other hand, places us in a good position to explain its capacities, powers, or functions – as well as all the other ‘sub-agential causes’ dealt with in this chapter. As suggested earlier, a definition of ‘animal’ in terms of capacities and powers, expresses a kind of functionalism: it presents an animal as a mover, a settler of matters, and an effector of

long-ranging changes in the physical world beyond its immediate self. These definitions are *prima facie* expressive of psychophysicality, and they reflect a kind of primitive truth about the animal's basic ability to move its own body. This crucial power is also the primary characteristic that makes us agents, and as far as explanations of our actions are concerned, the actualization of this power represents a kind of causation that is explanatorily primary to all other causal types (event, fact, property, and part) potentially involved in agential explanation.

Conclusion

In this thesis, I have attempted to strengthen the plausibility of the case for a non-causalist approach to the concept of action. The view that I have undertaken to defend is inspired by certain Aristotelian themes, and developed from one broadly understood variant of the disjunctivist idea. Indeed, throughout this thesis, I have tried to establish that certain Aristotelian themes are entirely compatible with the contemporary disjunctivist approach. The concluding remarks on the concept of an animal as integrating psychological and physical attributes suggest just one way in which this Aristotelian disjunctivism of action might be achieved.

It is hardly necessary to say that my investigation has not been exhaustive. Certainly there are many other possible disjunctivist articulations of agency, and there are just as many possible non-causalist approaches to consider. There is surely a significant amount of work that can be done on the variety of top-down causation considered and its potential accommodation of the notion of property fusion. There is also more to be said about whether the rapprochement between Aristotle's thought and contemporary issues is generally successful. Even if it is not, my hope has been that this latter conclusion is likewise a useful one, if only as a consideration of the deep problems that exist at the outset for any Aristotelian wishing (sometimes rather nostalgically, for better or for worse) to revitalize Aristotelian concepts within contemporary philosophy and its stance towards scientific compatibility.

My primary aims have been to defend the concept of action and agency against two particular tendencies in the contemporary literature: first, a certain post-Cartesian framework of understanding 'action', and second, the body of assumptions typically drawn from the canons of reductive materialism for application to action theory. With regard to these latter two aims, I hope that I have at least been able to establish that there *is* an intelligible alternative available.

Bibliography

- i. Historical Sources and Secondary Literature
- ii. Philosophy of Action and Perception
- iii. Metaphysics of Mind and Scientific Literature

i. Historical Sources and Secondary Literature

Aquinas. *Summa Theologiae* and Commentaries on *De Anima*, *Metaphysics*, *Physics*, and *Nicomachean Ethics*.

Aristotle. *Categories*.

— *De Anima*.

— *De Generatione Animalium*.

— *De Motu Animalium*.

— *Eudemian Ethics*.

— *Metaphysics*.

— *Nicomachean Ethics*.

— *Physics*.

— *Rhetoric*.

Broadie, Sarah. *Ethics with Aristotle* (1991). New York: Oxford University Press.

Burnyeat, Myles. ‘How Much Happens When Aristotle Sees Red and Hears Middle C?’ (1992). In Martha C. Nussbaum and Amelie Rorty (eds.), *Essays on Aristotle’s de Anima*. 421-34. New York: Clarendon Press.

Charles, David. *Aristotle’s Philosophy of Action* (1984). London: Duckworth.

- ‘Aristotle’s Psychological Theory’ (2009). *Proceedings of the Boston Area Colloquium in Ancient Philosophy, 2008, Volume 24*, ed. by J. J. Cleary and Gary M. Gurtler.
 - ‘Desire in Action: Aristotle’s Move’ (2011). In *Moral Psychology and Human Action in Aristotle*. Eds. Michael Pakuluk and Giles Pearson. New York: Oxford University Press.
 - ‘The Eudemian Ethics on the “Voluntary”’ (2012). In Fiona Leigh (ed.), *The Eudemian Ethics on the Voluntary, Friendship, and Luck: The Sixth S. V. Keeling Colloquium in Ancient Philosophy*.
- Coope, Ursula. ‘Aristotle on Action’ (2007). *Proceedings of the Aristotelian Society Supplementary Volume LXXXI*.
- ‘Change and its Relation to Potentiality and Actuality’ (forthcoming). In G. Anagnostopoulos (ed.), *Blackwell Companion to Aristotle*. Singapore: Blackwell.
- Descartes, René. ‘Letter to More’ (1649) In John Cottingham, Robert Stoothoff, and Dugald Murdoch (trans.), *The Philosophical Writings of Descartes* (1991). New York: Cambridge University Press.
- *Principles of Philosophy* (1644), trans. by R. P. Miller (1982). Netherlands: Springer.
 - *Meditations on First Philosophy* (1641), trans. by Donald Cress (1998). Indianapolis: Hackett.
 - *Discourse on Method* (1637), trans. by Donald Cress (1998). Indianapolis: Hackett.
- Heinaman, R. ‘Aristotle on Housebuilding’ (1985). *History of Philosophy Quarterly*, 2, 145-62.
- Hume, David. *An Enquiry Concerning Human Understanding* (1748), ed. Eric Steinberg (1993). Indiana: Hackett.

- *A Treatise of Human Nature* (1739), ed. David Norton (2000). New York: Oxford University Press.
- Johansen, T. K. *Aristotle on the Sense-Organs* (2007). Cambridge: Cambridge University Press.
- Kosman, L. A. ‘Aristotle’s Definition of Motion’ (1969). *Phronesis* 14, 40-62.
- Kostman, J. ‘Aristotle’s Definition of Change’ (1987). *History of Philosophy Quarterly*.
- Menn, Stephen. ‘Aristotle’s Definition of Soul and the Programme of the De Anima’ (2002). *Oxford Studies in Ancient Philosophy*, 22: 83-139.
- Nussbaum, Martha and Putnam, Hilary. ‘Changing Aristotle’s Mind’ (1995). In Nussbaum, M. and Rorty, A. O. (eds.), *Essays on Aristotle’s De Anima*. New York: Oxford University Press.
- Pakaluk, Michael and Pearson, Giles (eds). *Moral Psychology and Human Action in Aristotle* (2011). New York: Oxford University Press.
- Plato. *Laws*.
- *Phaedrus*.
- *Timaeus*.
- Shields, Christopher. ‘Soul as Subject in Aristotle’s *De Anima*’ (1988). *The Classical Quarterly*. Vol. 38:1,140-149.
- Waterlow, Sarah. *Nature, Change and Agency in Aristotle’s Physics* (1982). Oxford: Clarendon Press.
- Whiting, Jennifer. ‘Locomotive Soul’ (2002). *Oxford Studies in Ancient Philosophy*.
- Wittgenstein, Ludwig. *Philosophical Investigations* (2001/1953), trans. by Anscombe, G.E.M. Oxford: Basil Blackwell.

ii. Philosophy of Action and Perception

Aguilar and Buckareff. *Causing Human Actions: New Perspectives on the Causal Theory of Action* (2010). Cambridge, MA: MIT Press.

Anscombe, Elizabeth. 'The Causation of Action' (1983). In C. Ginet, ed., *Knowledge and Mind*. New York: Oxford University Press.

— 'Under a Description' (1979). In *The Collected Papers*, ii: *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell.

— *Intention* (2000/1957). Cambridge, MA: Harvard University Press.

Armstrong, David M. 'The headless woman illusion and the defence of materialism' (1968). *Analysis*, 29, 2: 48-9.

Belnap, Nuel; Perloff, Michael; Xu, Ming. *Facing the Future: Agents and Choices in Our Indeterminist World* (2001). New York: Oxford University Press.

Bishop, John. 'Agent-Causation' (1983). *Mind*, 92: 61-79.

Bonjour, Lawrence. *Epistemology: Classic Problems and Contemporary Responses* (2002). Lanham: Rowman and Littlefield.

Bratman, Michael. 'Two Faces of Intention' (1984). *The Philosophical Review*, 93, 3: 375-405.

— 'Reflection, Planning and Temporally Extended Agency' (2000). *Philosophical Review*, 109: 35-61.

— 'Two Problems About Human Agency' (2001). *Proceedings of the Aristotelian Society*, 101: 309-326.

- Brentano, Franz. *Psychology from an Empirical Standpoint* (1874). Ed. by Oskar Kraus, trans. by Rancurello, Terrell, and McAlister (1995). New York: Routledge.
- Burge, Tyler. ‘Primitive Agency and Natural Norms’ (2009). *Philosophy and Phenomenological Research*, 129, 2: 251-278.
- Byrne, Alex and Logue, Heather. *Disjunctivism: Contemporary Readings* (2008). Cambridge, MA: MIT Press.
- Chisholm, Roderick. ‘The Agent As Cause’ (1976). In Myles Brand and Douglas Walton (eds.), *Action Theory*. Boston: Reidel Publishing.
- Collins, Arthur. ‘The Psychological Reality of Reasons’ (1997). *Ratio*, 10, 2: 108-123.
- ‘Action, Causality, and Teleological Explanation’ (1984). In P. French et. al. (eds.), *Midwest Studies in Philosophy IX*. Minneapolis: University of Minnesota Press.
- ‘Teleological Reasoning’ (1978). *Journal of Philosophy*, 75: 540-550.
- Dancy, Jonathan. ‘Arguments From Illusion’ (1995). *The Philosophical Quarterly*, 45, 181: 421-438.
- *Practical Reality* (2002). New York: Oxford University Press.
- Danto, Arthur. *Analytical Philosophy of Action* (1973). New York: Cambridge University Press.
- With Sidney Morgenbesser, ‘What We Can Do’ (1963). *The Journal of Philosophy*, 60, 15: 435-445.
- ‘Basic Actions’ (1965). *American Philosophical Quarterly*, 2, 2: 141-148.
- Davidson, Donald. ‘Actions, Reasons and Causes’ (1963). *Journal of Philosophy*, 60: 685-700.

- ‘Agency’ (1971). In Robert Binkley, Richard Bronaugh, and Ausonia Marras (eds.), *Agent, Action and Reason*. Toronto: University of Toronto Press.
- Essays on Actions and Events* (1980). New York: Oxford University Press.
- ‘Rational Animals’ (1985). In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. New York: Basil Blackwell.
- ‘Thought and Talk’ (2001). In *Inquiries into Truth and Interpretation (Philosophical Essays of Donald Davidson)*. New York: Oxford University Press.
- *Problems of Rationality* (2004). New York: Clarendon Press.
- Donagan, Alan. *Choice: The Essential Element in Human Action* (1987). London: Routledge & Kegan Paul.
- Eng, Berent. *How We Act: Causes, Reasons, and Intentions* (2006). New York: Oxford University Press.
- Falvey, Kevin. ‘Knowledge in Intention’ (2000). *Philosophical Studies*, 99, 1: 21-44.
- Foot, Philippa. *Natural Goodness* (2001). New York: Oxford University Press.
- Ford, Anton; Hornsby, Jennifer; and Stoutland, Frederick, eds. *Essays on Anscombe’s Intention* (2011). Cambridge, MA: Harvard University Press.
- Frankfurt, Harry. ‘The Problem of Action’ (1978). In Mele (ed.), *The Philosophy of Action*. New York: Oxford University Press.
- Geach, Peter. *Mental Acts* (1957). London: Routledge.
- Gendler, Tamar and Hawthorne, John, eds. *Perceptual Experience* (2006). New York: Oxford University Press.
- Ginet, Carl. *On Action* (1990). Cambridge: Cambridge University Press.

- Grice, Paul. 'The Causal Theory of Perception' (1961). *Proceedings of the Aristotelian Society (Supplementary Volume)*, 35: 121-52.
- Haddock, Adrian. 'At One With Our Actions, But At Two With Our Bodies: Hornsby's Account of Action' (2005). *Philosophical Explorations*, 8, 2: 157-72.
- Haddock, Adrian and Macpherson, Fiona, eds. *Disjunctivism: Perception, Action, Knowledge* (2011). Oxford: Oxford University Press.
- Haldane, John. 'Identifying Privative Causes' (2011). *Analysis* 71, 4: 611-619.
- 'Is Every Action Morally Significant?' (2011). *Philosophy* 86, 3: 375-404.
 - 'A Return to Form in the Philosophy of Mind' (1998). *Ratio* XI, 3: 253-277.
 - 'On Some Metaphysical Presuppositions of Action' (1994). *The Heythrop Journal*, Vol. 35, 3: 296-303.
- Hampshire, Stuart. *Thought and Action* (1959). Notre Dame: University of Notre Dame Press.
- Hershfield, Jeffrey. 'Structural Causation and Psychological Explanation' (2001), *The Journal of Mind and Behavior*, Volume 22, 3: 249-261.
- Hornsby, Jennifer. *Actions* (1980). London: Routledge & Kegan Paul.
- 'Agency and Actions' (2004). In H. Steward and J. Hyman (eds.), *Agency and Action*. Cambridge: Cambridge University Press.
 - 'Alienated Agents' (2004). In M. De Caro and D. Macarthur (eds.), *Naturalism in Question*. Cambridge, MA: Harvard University Press.
 - and Stanley, Jason. 'Semantic Knowledge and Practical Knowledge' (2005). *Aristotelian Society Supplementary Volume*, 79, 1: 107-130.

- ‘Knowledge and Abilities in Action’ (2007). In *Proceedings of the 29th International Ludwig Wittgenstein Symposium*.
 - ‘Actions in Their Circumstances’ (2011). In Anton Ford, Jennifer Hornsby, and Frederick Stoutland (eds.), *Essays on Anscombe’s Intention*. Cambridge, MA: Harvard University Press.
- Humberstone, I. L. ‘Direction of Fit’ (1992). *Mind*, Vol. 101, 401: 59-83.
- Hursthouse, Rosalind. ‘Arational Actions’ (1991). *The Journal of Philosophy*, Vol. 88, 2: 57-68.
- Juarerro, Alicia. *Dynamics in Action: Intentional Behavior as a Complex System* (1999). Cambridge, MA: MIT Press.
- Lavin, Douglas. ‘Must There Be Basic Action?’ (2012). *Noûs* 00, 0: 1-32.
- Lowe, E.J. *Personal Agency: The Metaphysics of Mind and Action* (2010). New York: Oxford University Press.
- Martin, M.G.F. ‘The Transparency of Experience’ (2002). *Mind and Language*, Vol. 4, 4: 376-425.
- ‘The Limits of Self-Awareness’ (2004). *Philosophical Studies*, 120: 37-89.
 - ‘On Being Alienated’ (2006). In T. Gendler, J. Hawthorne (eds.), *Perceptual Experience*. New York: Oxford University Press.
- McDowell, John. ‘Non-Cognitivism and Rule-Following’ (1981). In S. Holtzman and Christopher M. Leich (eds.), *Wittgenstein: To Follow a Rule*. London: Routledge.
- ‘Singular Thought and the Extent of Inner Space’ (1986). In P. Petit and J. McDowell (eds.), *Subject, Thought, and Context*. Oxford: Clarendon Press.
- Mind and World* (1996). Cambridge, MA: Harvard University Press.

- Mele, Alfred. *Springs of Action* (1992). New York: Oxford University Press.
- *Motivation and Agency* (2003). New York: Oxford University Press.
 - *Effective Intentions: The Power of Conscious Will* (2009). New York: Oxford University Press.
- Moran, Richard. ‘Anscombe on “Practical Knowledge”’ (2004). *Royal Institute of Philosophy Supplement*, 55: 43-68.
- O’Shaughnessy, Brian. ‘Trying (As the Mental ‘Pineal Gland’)’ (1973). *Journal of Philosophy*, 70: 365-386.
- *The Will: A Dual-Aspect Theory*, Vol. I-II (1980). New York: Cambridge University Press.
- Peacocke, Christopher. *Holistic Explanation: Action, Space, Interpretation* (1979). Oxford: Clarendon Press.
- Putnam, Hilary. ‘The Meaning of “Meaning”’ (1975). *Minnesota Studies in the Philosophy of Science*, 7: 131-193.
- Robinson, Howard. *Perception* (1994). London: Routledge.
- Rödl, Sebastian. ‘Two Forms of Practical Knowledge and Their Unity’ (2011). In A. Ford, J. Hornsby, and F. Stoutland (eds.), *Essays on Anscombe’s Intention*. Cambridge, MA: Harvard University Press.
- Schlosser, M.E. ‘Basic deviance reconsidered’ (2007). *Analysis*, 67: 186-94.
- ‘Agency, Ownership, and the Standard Theory’ (2012). In Buckareff, Aguilar, and Frankish (eds.), *New Waves in Philosophy of Action*. New York: Macmillan.
- Searle, John. *Intentionality in Action* (2001). Cambridge, MA: MIT Press.

- Sehon, Scott R. ‘Deviant Causal Chains and the Irreducibility of Teleological Explanation’ (1997). *The Pacific Philosophical Quarterly*, 78: 195-213.
- Teleological Realism: Mind, Agency, and Explanation* (2005). Cambridge, MA: MIT Press.
- Setiya, Kieran. ‘Explaining Action’ (2003). *Philosophical Review* 112, 3: 339-393.
- ‘Practical Knowledge’ (2008). *Ethics*, 118, 3: 388-409.
- Smith, Michael. ‘The Humean Theory of Motivation’ (1987). *Mind*, 96: 36-61.
- ‘The Structure of Orthonomy’ (2004). *Royal Institute of Philosophy Supplement*, 55: 165-193.
- ‘Four Objections to the Standard Story of Action (and Four Replies)’ (2012). *Philosophical Issues*, 22: 387-401.
- Snowdon, Paul. ‘Perception, Vision, and Causation’ (1980). *Proceedings of the Aristotelian Society*. 81: 175-92.
- Stout, Rowland. *Things That Happen Because They Should: A Teleological Approach to Action* (1996). New York: Oxford University Press.
- *Action* (2005). Trowbridge: Acumen Publishing.
- Stoutland, Frederick. ‘Introduction: Anscombe’s *Intention* in Context’ (2011). In Anton Ford, Jennifer Hornsby, and Frederick Stoutland (eds.), *Essays on Anscombe’s Intention*. Cambridge, MA: Harvard University Press.
- Strawson, P.F. ‘Perception and Its Objects’ (1979). In G.F. Macdonald (ed.), *Perception and Identity: Essays Presented to A.J. Ayer with His Replies to Them*. London: Macmillan.
- Tännsjö, Torbjörn. ‘On deviant causal chains – no need for a general criterion’ (2009).

Analysis 69, 3: 469-473.

Thompson, Michael. *Life and Action: Elementary Structures of Practice and Practical Thought* (2008). Cambridge, MA: Harvard University Press.

‘Anscombe’s Intention and Practical Knowledge’ (2011). In Ford, Anton, Hornsby, Jennifer, and Stoutland, Frederick (eds.), *Essays on Anscombe’s Intention*. Cambridge: Harvard University Press.

Velleman, David. ‘What Happens When Someone Acts?’ (1992). *Mind*: Vol. 101, 403: 461-481.

Wilson, George. *The Intentionality of Human Action* (1989). Redwood City: Stanford University Press.

iii. Metaphysics of Mind and Scientific Literature

Ayers, M. *The Refutation of Determinism* (1968). London: Methuen.

Bickle, John. ‘Reducing mind to molecular pathways: explicating the reductionism implicit in current cellular and molecular neuroscience’ (2006). *Synthese*, 151: 411-434.

Broadie, Sarah. ‘Agency and Determinism in A Metaphysics for Freedom’ (2013). *Inquiry: An interdisciplinary Journal of Philosophy*, 56, 6: 571-582.

Campbell and Bickhard, ‘Physicalism, Emergence, and Downward Causation’ (2011). *Axiomathes*, 21: 33-56.

Carpenter, R.H.S. *Neurophysiology*, 4th Ed. (2003). London: Arnold.

- Cartwright, Nancy. 'Causation: One Word, Many Things' (2004). *Philosophy of Science*, 71, 5: 805-819.
- Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory* (1996). Oxford: Oxford University Press.
- 'What is a neural correlate of consciousness?' (2000). In Metzinger, Thomas (ed.), *Neural Correlates of Consciousness*. Cambridge, MA: MIT Press.
- 'Strong and Weak Emergence' (2006). In Clayton, P. and Davies, P. (eds.), *The Re-emergence of Emergence*. New York: Oxford University Press.
- Chisholm, Roderick. 'Human Freedom and the Self' (1964). The Lindley Lecture, published by the Department of Philosophy of the University of Kansas.
- Cohen, Marc. 'Hylomorphism and Functionalism' (1992). In M. Nussbaum and A. Rorty (eds.), *Essays on Aristotle's De Anima*. New York: Clarendon.
- Crane, Tim and Mellor, D. H. 'There is No Question of Physicalism' (1990). *Mind* 99, 394: 185-206.
- Dretske, Fred. *Explaining Behaviour* (1988). *Reasons in a World of Causes*. Cambridge, MA: MIT Press.
- Farrer, Austin. *The Freedom of the Will* (1958). London: A&C Black.
- Fine, Kit. 'Acts, Events, and Things' (1982). In *Proceedings of the International Wittgenstein Symposium* (Leinfellner et al: Language and Ontology).
- 'Compounds and Aggregates' (1994). *Noûs* 28, 2: 137-158.
- Fodor, Jerry. 'Making Mind Matter More' (1989). *Philosophical Topics* 67: 59-79.
- Gaser, Christian, and Schlaug, Gottfried. 'Gray Matter Differences between Musicians and Nonmusicians' (2003). *Annals of the New York Academy of Sciences* 999, 1: 514-

517.

- Godfrey-Smith, Peter. 'Causal Pluralism' (2010). In H. Beebe, C. Hitchcock, and P. Menzies, (eds.), *Oxford Handbook of Causation*. New York: Oxford University Press.
- Haggard, P. 'Human Volition: Towards a Neuroscience of Will' (2008). *Nature Reviews Neuroscience*, 9: 934-946.
- Hanna, Robert and Maiese, Michelle. *Embodied Minds in Action* (2009). New York: Oxford University Press.
- Humphreys, Paul. 'Aspects of Emergence' (1996). *Philosophical Topics* 24: 53-70.
- 'How Properties Emerge' (1997a). *Philosophy of Science* 64, 1: 1-17.
- 'Emergence, Not Supervenience' (1997b). *Philosophy of Science* 64: 337-345.
- Hurley, Susan and Noë, Alva. 'Neural Plasticity and Consciousness' (2003). *Biology and Philosophy* 18, 1: 131-168.
- Jaworski, William. 'Hylomorphism and Mental Causation' (2006a). *Proceedings of the American Catholic Philosophical Association*, 79: 201-216.
- 'Hylomorphism and Post-Cartesian Philosophy of Mind' (2006b). *Proceedings of the American Catholic Philosophical Association*, 80: 209-224.
- Kim, Jaegwon. "'Strong" and "Global" Supervenience Revisited' (1987). *Philosophy and Phenomenological Research* 48, 2: 315-326.
- 'The Myth of Nonreductive Materialism' (1989). *Proceedings and Addresses of the American Philosophical Association*, 63, 3: 31-47
- *Supervenience and Mind* (1993). New York: Cambridge University Press.
- 'Making Sense of Emergence' (1999). *Philosophical Studies*, 95: 3-36.

- ‘Blocking Causal Drainage and Other Maintenance Chores with Mental Causation’ (2003). *Philosophy and Phenomenological Research* 67, 1: 151-176.
- *Physicalism, Or Something Near Enough* (2005). Princeton: Princeton University Press.
- ‘Being Realistic about Emergence’ (2006). In Clayton, P. and Davids, P. S. (eds.), *The Re-Emergence of Emergence*. New York: Oxford University Press.

Lakoff, George and Johnson, Mark. *Philosophy in the Flesh: the Embodied Mind and its Challenge to Western Thought* (1999). New York: Basic Books.

Libet et. al. ‘Time of Conscious Intention to Act In Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act’ (1983). *Brain*, 106: 623-642.

Loux, Michael. ‘Aristotle on Matter, Form, and Ontological Strategy’ (2005). *Ancient Philosophy* 25, 1: 81-123.

Lowe, E.J. ‘Substance Causation, Powers, and Human Agency’ (2013). In S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson (eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press.

- *An Introduction to the Philosophy of Mind* (2000). Cambridge: Cambridge University Press.

Macdonald, Cynthia and Macdonald, Graham. ‘Emergence and Downward Causation’ (2010). In Macdonald C. and Macdonald G. (eds.), *Emergence in Mind*. New York: Oxford University Press.

Marcus, Eric. ‘Mental Causation in a Physical World’ (2005). *Philosophical Studies* 122: 27-50.

Matsushashi, M. and Hallett, M. ‘The Timing of the Conscious Intention to Move’ (2008). *European Journal of Neuroscience*, 28: 2344-2351.

- McCulloch, Warren and Pitts, Walter. 'A logical calculus of the ideas immanent in nervous activity' (1943). *Bulletin of Mathematical Biophysics* 5: 115-133.
- Merricks, Trenton. 'Composition as Identity, Mereological Essentialism, and Counterpart Theory' (1999). *Australasian Journal of Philosophy*. 77, 2: 192-195.
- Noë, Alva and Thompson, Evan. 'Are There Neural Correlates of Consciousness?' (2004). *Journal of Consciousness Studies*, 11, 1: 3-28.
- O'Connor, Timothy. *Persons and Causes: The Metaphysics of Free Will* (2000). New York: Oxford University Press.
- Pascual-Leone, Alvaro. 'The Brain That Plays Music and Is Changed by It' (2001). *Annals of the New York Academy of Sciences*, 930, 1: 315-329.
- Quine, W. V. *The Ways of Paradox and Other Essays* (1966). Cambridge, MA: Harvard University Press.
- Rauschecker, Josef P. 'Cortical Plasticity and Music' (2001). *Annals of the New York Academy of Sciences*, 930, 1: 330-336.
- Ryle, Gilbert. *The Concept of Mind* (1949). London: Hutchinson.
- Scaltsas, Theodore. 'Is a Whole Identical to its Parts?' (1990). *Mind*, 99, 396: 583-598.
- 'Substantial Holism' (1994). In T. Scaltsas, D. Charles, and M. L. Gill (eds.), *Unity, Identity, and Explanation in Aristotle's Metaphysics*. Oxford: Oxford University Press.
- Schlaug, Gottfried. 'The Brain of Musicians: A Model for Functional and Structural Adaptation' (2001). *Annals of the New York Academy of Sciences*, 930, 1: 281-299.
- Simons, Peter. *Parts: A Study in Ontology* (1987). Oxford: Clarendon Press.

- Smyth, Mary and Wing, Alan. *The Psychology of Human Movement* (1984). London: Academic Press.
- Stein, J.F. and Stoodley, C.J. *Neuroscience: An Introduction* (2006). London: John Wiley and Sons.
- Steward, Helen. *A Metaphysics for Freedom* (2012). New York: Oxford University Press.
- Strawson, P. F. *Analysis and Metaphysics: An Introduction to Philosophy* (1992). New York: Oxford University Press.
- *Individuals: An Essay in Descriptive Metaphysics* (1959). London: Routledge.
- Taylor, A. *Plato: The Man and his Work* (2001/1927). New York: Dover.
- Trevena, Judy and Miller, Jeff. ‘Brain preparation before a voluntary action: Evidence against unconscious movement initiation’ (2010). *Consciousness and Cognition* 19: 447-456.
- Varela, Francisco J., Thompson, Evan, and Rosch, Eleanor. *The Embodied Mind* (1991). Cambridge, MA: MIT Press.
- Wegner, Daniel. *The Illusion of Conscious Will* (2002). Cambridge, MA: MIT Press.
- ‘The Mind’s Best Trick: How We Experience Conscious Will’ (2003). *Trends in Cognitive Sciences*, 7, 2: 65-69.
- ‘Apparent Mental Causation: Sources of the Experience of Will’ (1999). *American Psychology*, 54, 7: 480-92.
- Wiggins, David. *Sameness and Substance Renewed* (2001). Cambridge: Cambridge University Press.
- Zatorre, Robert J. ‘Music and the Brain’ (2003). *Annals of the New York Academy of Sciences*, 999, 1: 4-14.