THE NATURE OF AGENTIVE AWARENESS

by

MYRTO MYLOPOULOS

A dissertation submitted to the Graduate Faculty in Philosophy
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York

2013

©2013

MYRTO MYLOPOULOS

All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Prof. Barbara Montero

Date — Chair of Examining Committee

Prof. Iakovos Vasiliou

Date — Executive Officer

Prof. David Rosenthal, Advisor

Prof. Jesse Prinz

Prof. Élisabeth Pacherie

Prof. Frédérique de Vignemont

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE NATURE OF AGENTIVE AWARENESS

by

Myrto Mylopoulos

Advisor: Professor David Rosenthal

The way in which we are subjectively aware of our passive movements stands in stark contrast to the way in which we are typically aware of our actions. Following Bayne and Pacherie (2007), I call the latter type of awareness *agentive awareness*. A robust literature on agentive awareness has recently emerged, but there is as of yet no consensus as to its underlying nature. The goal of this dissertation is to give a complete account of agentive awareness that is sensitive to a range of theoretical and empirical considerations.

There are three main questions that frame theorizing surrounding agentive awareness. The first is the question of what kind of awareness is agentive awareness—in other words, in virtue of what kind of mental state (e.g., sensory, cognitive) is one aware of oneself as acting? In Chapter 2, I argue that no sensory approach to agentive awareness is workable. In particular, I argue that if one is agentively aware in virtue of being in suitable sensory states, then such states must be the outputs of some sensory modality. But there is no sensory modality within which to locate these states.

Second, there is the question of how agentive awareness relates to action control. Is it a function of low-level, sensorimotor control, as some have argued? Or high-level, intentional control, as others maintain? In Chapter 3, I argue against dominant low-level accounts of agentive awareness that are pitched in terms of a popular neurocomputational model of sensorimotor control developed by Chris Frith: the comparator model. I

evaluate the empirical case for this approach, and argue that it fails to support it. Moreover, I argue that there are dissociations between sensorimotor control and agentive awareness that raise doubts about the success of any low-level account of agentive awareness.

A third question pertains to the psychological mechanisms underlying agentive awareness. What events must take place at the psychological level in order for agentive awareness to arise? There is a broad consensus among theorists that agentive awareness arises out of a matching process between our intentions and our actions. The most influential version of this view has been championed by Daniel Wegner. In Chapter 4, I argue against Wegner's view, and matching accounts in general, on the grounds that (i) the empirical evidence cited in their favor does not, in fact, hold up, and (ii) they are not sensitive to the reliable character of our intentions, and (iii) there are cases in which agentive awareness arises in the absence of a match between an intention and an action.

These considerations point in the end to an account of agentive awareness on which it is non-sensory, located at the level of intentional control, and does not require a match between an intention and an action to arise. In Chapter 5, I develop a novel account of agentive awareness along these lines, arguing that one is agentively aware in virtue of being in suitable cognitive states, i.e., thoughts, which are formed on the basis of executive intentions, i.e., intentions to do something here and now. This account does justice to pre-theoretical desiderata, as laid out in Chapter 1, avoids the pitfalls by which other accounts are hindered, and enjoys ample empirical support.

Acknowledgements

Parts of this dissertation were presented at the Association for the Scientific Study of Consciousness, City College, CUNY, the CUNY Cognitive Science Symposium, the European Society for Philosophy and Psychology, New York University, Simon Fraser University, the Southern Society for Philosophy and Psychology, the Society for Philosophy and Psychology, Stanford University, the University of Edinburgh, and the University of Western Ontario. I thank these audiences for their helpful discussion of this material.

This dissertation has greatly benefitted from the valuable feedback and kind support of a number of colleagues and professors. I am grateful to the following people: Mark Alfano, Tim Bayne, Todd Beattie, Salas Sanchez Bennasar, Ned Block, Maria Brincker, Richard Brown, Michael Brownstein, Wesley Buckwalter, Glenn Carruthers, Jennifer Corns, Josh Dulberger, Kyle Ferguson, Lily Frank, Ellen Fridland, Chris Frith, Daniel Harris, Grace Helton, Marcela Herdova, Linus Huang, Zoe Jenkin, Alex Kiefer, Michal Klincewicz, Jean-David Lafrance, Kirstie Laird, Hakwan Lau, Pete Mandik, Tony Marcel, Al Mele, David Pereplyotchik, Jake Quilty-Dunn, Brian Robinson, Lee-Anna Sangster, Aaron Schurger, John Schwenkler, Dan Shargel, Joshua Shepherd, Henry Shevlin, Chris Alen Sula, Jesse Summers, Jennifer Warriner, and Benjamin Young. I am especially grateful to Jacob Berger for his constant support and friendship.

I wish to thank Tony Ro and the members of his Cognitive Neuroscience lab at CCNY for welcoming me there.

My very many thanks to the members of my dissertation committee, Barbara Montero, Élisabeth Pacherie, Jesse Prinz, and Frédérique de Vignemont, for their insightful feedback, support, and guidance throughout the writing of this dissertation.

I am enormously grateful to my advisor, David Rosenthal, for his tireless support, warm encouragement, solid guidance, and the great deal that he has taught me over the years. I am especially grateful to him for a certain phone conversation.

I wish to express my deep gratitude to my friends—both near and far—and to my family, John Mylopoulos, Chryssafi Mylopoulos, Maria Mylopoulos, and Jon Newman, for their love and support.

I would not have been able to complete this project without the unwavering support, love, patience, and humor of Laura Jayne Martin. Thank you.

Table of Contents

List of Figures

CHAPTER 1: THE THEORETICAL LANDSCAPE

§1 Introduction

Consider the following vignette, with an eye to the types of subjective experience that the agent in question is having:

> This morning I was awakened by the sound of someone practising the violin. I dozed a bit, then got up, washed, shaved, dressed, and went downstairs, turning off the light in the hall as I passed. I poured myself some coffee, stumbling on the edge of the dining room rug, and spilled my coffee fumbling for the New York Times. (Davidson, 1971/2001b, p. 43)

There is a clear difference between the way in which we are typically subjectively aware of events like getting up, washing, walking, and pouring coffee, and the way in which we are typically subjectively aware of events like stumbling, fumbling, and spilling. In the latter cases, we feel that we merely observe these events unfolding outside of our control. When we get up, wash, walk, and pour coffee, however, we are aware of these events *as our actions*—as expressions of our own agency. Following Bayne & Pacherie (2007), I will refer to this type of awareness as *agentive awareness*.[1]

Historically, philosophers have sometimes gestured towards agentive awareness and related experiences. For example, Hume (1739/2000) wrote of the will as "the

[1] Agentive awareness has gone under various alternative names in the literature, most popularly "the sense of agency" (e.g., Gallagher, 2000; Marcel, 2003). Other labels include the "phenomenology of agency" (Pacherie, 2008), "feeling of doing" , "experience of authorship" (Prinz, 2012), "action consciousness", (Prinz, 2007) and "control consciousness" (Mandik, 2010). I prefer the term 'agentive awareness' because it has a theoretically neutral character—leaving it open, for example, as to whether agentive awareness is properly understood as a 'sense', a 'feeling' or an 'experience'.

internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body or new perception of our mind" (III, i, 1). While ostensibly making these remarks about the will itself, Hume also seems to posit an experience that accompanies our intentional actions, which many would view as a feature of agentive awareness. And Locke (1689/1975) observed, in his *Essay Concerning Human Understanding*, that, "... we find by experience that barely by willing it, barely by a thought of the Mind, we can move the parts of our bodies which were before at rest" (II, xxi, 1). Here Locke is referring to the sense that we have of being the agents of our actions—a central feature of agentive awareness.

Other than these occasional nods, however, up until relatively recently, theorists have largely ignored agentive awareness. They have chosen instead to focus their research on other aspects of our conscious mental lives. This raises the question: Why have theorists started to pay attention to agentive awareness only now?

One likely explanation is the fascinating body of research coming out of cognitive science, which presents stimulating challenges regarding the veridicality and reliability of agentive awareness and related experiences.[2] A classic challenge comes from the famous work of the neuroscientist Benjamin Libet in the 1980s, who purported to show that our conscious decisions to act arrive too late on the scene to cause our actions—they are caused instead by nonconscious events in the brain that precede our conscious decisions by about one third of a second. As Libet (1985) provocatively summarized his findings, "the brain evidenctly 'decides' to initiate, or, at the least, prepare to initiate the act at a

[2] Also very influential has been the work of Frith (1992) and Jeannerod (2006) who offer robust theoretical frameworks for understanding the psychology and neuroscience of action and its accompanying disorders.

time before there is any subjective awareness that such a decision has taken place" (p. 536).

Even more recently, another striking claim comes from the psychologist Daniel Wegner's (2002) influential book, *The Illusion of Conscious Will.* Wegner singlehandedly did much to galvanize the recent interest in agentive awareness by championing the view that agentive awareness, or 'conscious will' as he calls it, is in important respects illusory. As he puts it:

> ... it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. Although it is sobering and ultimately accurate to call all this an illusion**,** it is a mistake to conclude that the illusion is trivial. On the contrary, the illusions piled atop apparent mental causation are the building blocks of human psychology and social life. (p. 342)

It is reasonable to suppose that, in light of these worrisome challenges, many have thought that a vindication of agentive awareness requires a proper understanding of it.

But why defend agentive awareness against these challenges in the first place? Put differently, why does it matter if Wegner and Libet are correct? Here is where, perhaps, the root of the interest in agentive awareness, from all angles, is revealed: the relationship it is thought to have to free will. For many, agentive awareness is a subjective indication of our own free will. It is a primary source of evidence that we have a certain type of control over our actions, in the way required for us to be free and morally responsible agents. Indeed, theories of free will, it is often urged, ought to answer to how our actions

subjectively *feel* to us—particularly aspects of how they feel that indicate that they are free or not free. For example, Nahmias et al. (2004) write that,

> [t]heories of free will are more plausible when they capture our intuitions and experiences than when they explain them away. Thus, philosophers generally want their theories of free will to aptly describe the experiences we have when we make choices and feel free and responsible for our actions. (p. 162)

If theories of free will must explain our phenomenology of agency, and if this phenomenology serves, in turn, as evidence in support of theories of free will, then this is motivation for understanding the nature of this phenomenology. Or so many seem to hold.[3]

I will not be further discussing the relationship between agentive awareness and free will here. My present aim is simply to offer a credible diagnosis of the surge of interest in the topic. But whatever the reason or combination of reasons for the newfound attention, a sophisticated literature is now happily emerging surrounding this phenomenon, in philosophy, psychology, and neuroscience alike, and a fertile theoretical landscape is finally starting to take shape.

What has not emerged, perhaps unsurprisingly, is a consensus regarding the nature of agentive awareness. In this dissertation, I will survey and critique competing accounts in the literature, and develop an account of my own that is sensitive to a range

[3] While I do think that this is a major reason for the recent explosion of interest in agentive awareness, I have my doubts that phenomenology of any sort has a role to play in theories of free will. For it is not clear what kinds of predictions these theories make, if any, concerning our phenomenology. Do libertarian theories, for example, predict that we feel as though we have the ability to do otherwise when we act? Arguably not (Mylopoulos & Lau, forthcoming).

of theoretical and empirical considerations. In this chapter, I will trace the main positions and questions that presently frame this burgeoning literature, and orient the reader for the remainder of the dissertation.

I start, in the next section, by cataloguing the main desiderata that a theory of agentive awareness should aim to satisfy.

§2 A Theory of Agentive Awareness: Desiderata

Loosely speaking, and as a first pass, agentive awareness may be understood as one's awareness of acting. But a moment's reflection reveals that this is far too crude a characterization. After all, one is sometimes aware of others as acting, and though there may be some overlap in the psychological mechanisms that allow us both to act and to observe the actions of others (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), agentive awareness refers exclusively to the awareness of one's *own* actions. First and foremost, agentive awareness is a form of *self-awareness*—it is awareness of oneself as acting. So any theory of agentive awareness must also be an account of self-awareness, at least of this particular type.

There are different ways in which one comes to be aware of oneself as acting. Suppose, for example, that I notice an irritated expression on my neighbor's face, leading me to become suddenly aware that I am tapping my foot mindlessly. My foot-tapping is an action of mine, but in this case, I come to be aware of it in an unusual way, that is, by consciously observing what I am doing after being tipped off by my neighbor's facial expression. This is *not* the way in which we typically come to be aware of our own actions. Danto and Morgenbesser (1963) nicely depict the relevant contrast:

> If one day I should notice that my arm was rising and lowering, and then realize that, if I had not noticed, I would not have known it was doing this, it would be for me a terrifying experience, a sign that I had lost contact with part of myself, that my arm had become an alien entity. (p. 441)

When it comes to awareness of our own actions, unlike the awareness we have of the actions of others, the awareness comes about in a *subjectively unmediated* way: it does not *seem* to rely on observation or inference. This characteristic property of agentive awareness has been frequently noted—sometimes under the label "knowledge without observation" (Anscombe, 1966) or "knowledge from the inside" (Eilan & Roessler, 2003). A theory of agentive awareness must account for this special feature of such awareness.

Agentive awareness is commonly thought to have a distinctive *phenomenal character* that cannot be reduced to the phenomenology of bodily movement. As Horgan, Tienson, and Graham (2003) put it:

> We maintain that there is 'something it is like' to behave in a way that constitutes voluntary action, something phenomenologically distinctive that incorporates but goes beyond the phenomenology of one's own bodily motion. (p. 323)

This conviction has been widely echoed in cognitive psychology and neuroscience as well. For example, around the same time, Haggard and Johnson (2003) write: "[V]oluntary actions, though generally neglected in psychology, have a distinctive phenomenology which can be studied both qualitatively and quantitatively" (p. 83).

Indeed, in line perhaps with what Haggard and Johnson (2003) have in mind as the "qualitative" approach to examining this phenomenology, there have been a number of attempts in the literature to capture the many components or facets of the phenomenology of agency. Horgan, Tienson, and Graham (2003) themselves focus on three aspects of the phenomenology: what they call "self as source", "purposiveness", and "voluntariness". Bayne and Levy (2006), by contrast, divide the phenomenology into what they call the "experience of mental causation", the "experience of authorship", and the "experience of effort". And Pacherie (2008) understands the phenomenology to involve the 'sense of intentional causation", the "sense of initiation", and the "sense of control".

Alongside these efforts to get a grip on the phenomenology of agency, it is simultaneously acknowledged that it has a "thin" and "evasive" character to it (e.g., Metzinger, 2006). It is a background experience, according to many theorists—not a salient part of conscious awareness. For example, Frith (2005) describes this unusual fringe quality of the experience as follows:

> Paradoxically it seems as if the mark of the self in action is that we have very little experience of it. […] We know that we are agents and that we are successfully causing the world to change. But as actors we move through the world like shadows glimpsed only occasionally from the corner of an eye. (p. 767)

A theory of agentive awareness, then, must give some account of the phenomenology of agency—explaining either why it has the features that theorists have attributed to it, or why, if it does not have these features, it has seemed to so many to have them.

At the practical level, the purportedly thin and evasive quality of the phenomenology of agency has meant that it is commonly picked out, at least in the first instance, in a contrastive manner, by describing what it is *not* like. For the purposes of drawing the relevant contrast, theorists have leaned heavily on the clinical literature, in which several pathological conditions relating to action and the awareness thereof have been identified. The most commonly discussed disorders, and the ones on which I will focus, are: (i) delusions of control in schizophrenia, (ii) anarchic hand syndrome, (iii) utilization behavior, and (iv) anosognosia for hemiplegia.

Delusions of control in schizophrenia are a subtype of what researchers working on this condition call *passivity experiences* (Frith, Blakemore, & Wolpert, 2000; Schneider, 1957). In general, passivity experiences involve the attribution of an action, mental event, or sensory experience to an external source. They include a range of phenomena, such as thought insertion ("Thoughts come into my mind from outer space"), somatic passivity experiences ("I have tingling feelings in my legs caused by electric currents from an alternator"), and made emotions ("It puts feelings into me: joy, happiness, embarrassment, depression. It just puts it in and I feel the glow spread over me") (all quotations from Frith, Blakemore, & Wolpert, 2000, p. 358).

In delusions of control, schizophrenic individuals experience their actions passively—attributing them to alien sources. Their reports are rather striking. For instance, one person suffering from such delusions reported that "[m]y grandfather hypnotized me and now he moves my foot up and down" (Frith et al., 2000). Another individual insisted that, "the force moved my lips. I began to speak. The words were made for me" (Mellors, 1970). Yet another patient, having just made an arm movement,

complained that: “I felt like an automaton, guided by a female spirit who had entered me during it” (Spence et al., 1997). Clearly, agentive awareness is lacking for many of the actions that such individuals experience.

Anarchic hand syndrome (AHS) also involves startling disruptions to agentive awareness.[4] The condition results from brain damage, often to the Supplementary Motor Area (SMA) or anterior corpus callosum (Della Sala, 2005), which leads individuals to perform complex, goal-oriented bodily movements, all the while denying authorship over them. Patients with anarchic hand syndrome are unable to inhibit the movements of their anarchic limb, except by indirect methods, such as using their other hand to stop it, and experience great distress and frustration at their condition (Biran, Giovannetti, Buxbaum, & Chatterjee, 2006). Sometimes their anarchic hand will even behave at cross-purposes with their well-functioning limb, as in the reported case of one individual whose anarchic hand unbuttoned his shirt, while his other hand was attempting to button it up (Bogen, 1985). Those with anarchic hand syndrome do not, however, like schizophrenic individuals with delusions of control, attribute their anarchic movements to external agents—though they might report that it is “as if” the limb has “a will of its own” (Biran et al., 2006; Marchetti & Della Sala, 1998).

The movements of the anarchic hand are often triggered by environmental stimuli, and in many cases, the hand plays out routine actions. For example, in the case of one individual, JC, his anarchic hand “reached for light switches, repeatedly pressed buttons

[4] Note that sometimes anarchic hand syndrome is erroneously referred to as ‘alien hand syndrome’ (e.g., Assal, Schwartz, & Vuilleumier, 2007). But the latter refers to a disruption in the *sense of ownership* over one’s limb—that is, individuals deny that the limb in question belongs to them. Those suffering from anarchic hand syndrome, by contrast, do not deny ownership of the relevant limb, but instead lack a sense of agency for its movements (see Marchetti & Della Sala, 1998).

on the television remote control, and groped for his left hand or face during sleep" (Giovannetti, Buxbaum, Biran, & Chatterjee, 2005, p. 77). In another particularly vivid case of the disorder, Banks et al. (1989) describe the following contradictory behaviors exhibited by one of their anarchic hand patients:

> While playing checkers on one occasion, the left hand made a move he did not wish to make, and he corrected the move with the right hand; however, the left hand, to the patient's frustration, repeated the false move. On other occasions, he turned the pages of the book with one hand while the other tried to close it; he shaved with the right hand while the left one unzipped his jacket; he tried to soap a washcloth while the left hand kept putting the soap back in the dish; and he tried to open a closet with the right hand while the left one closed it. (p. 457)

Plainly, these individuals do not have agentive awareness over these actions—and a complete theory of agentive awareness should have some insight to offer as to why.

A third type of commonly discussed disruption to agentive awareness is utilization behavior (UB). Those with this condition, exhibiting damage to the pre-SMA—the anterior portion of the SMA—show repetitive, routine behavior that is heavily reactive to external stimuli. One patient, for instance, donned one pair of glasses after another as they were placed in front of him. Another drank several glasses of water consecutively, well beyond the point of thirst (Lhermitte, 1983). Utilization behavior is similar to the behavior associated with anarchic hand syndrome, both involving as they do highly stimulus-driven movements. But the attitude that those with utilization behavior take towards their movements is markedly different: they do not report distress or surprise at what they are doing, and often confabulate reasons if asked to explain their

behavior. Indeed, one individual is reported as having utilization behavior afflicting his right hand and anarchic hand syndrome affecting the left—he was disturbed by the behavior of the latter, but not the former (Marcel, 2003; see also Bayne & Pacherie, 2007 for discussion of the differences between anarchic hand syndrome and utilization behavior). This condition, too, begs for an explanation.

Fourth, consider anosognosia for hemiplegia (AHP), a disorder typically involving right-brain damage, and characterized by a lack of awareness of paralysis of the contralesional side of the body (Fotopoulou et al., 2008). The disorder is heterogeneous, manifesting itself in various ways, with some individuals denying paralysis completely, e.g., "Of course, I can use my left arm. How do you think I scratch my right arm when it is itchy during the night?" (Fotopoulou et al., 2008, p. 5), others acknowledging paralysis, but downplaying it significantly, and others verbally reporting that they are paralyzed, but nonetheless attempting to perform actions with their paralyzed limb (Marcel, Tegner, & Nimmo-Smith, 2004). Many of these patients report performing actions with their paralyzed limbs while they remain motionless. And it is arguable that, at least some of the time, they are agentively aware when they do so. So this is another type of case that must be accounted for by a theory of agentive awareness.

The verdict is still out on the underlying nature of these mystifying disorders—there is much that we do not yet understand. Still, a theory of agentive awareness should be able to venture an explanation of why agentive awareness is missing in these perplexing cases. Perhaps these explanations cannot be reinforced until much further down the road when more questions have been settled. But any theory of agentive

awareness that has nothing to say regarding these so-called *disorders of volition* (Sebanz & Prinz, 2006) is at a serious disadvantage compared with those theories that do.

A final aspect of agentive awareness that must be accounted for in a complete theory is its *representational character*. Agentive awareness is taken to have representational content—it represents the world as being a certain way, that is, it represents one's own agency. There is a question about how "thick" or "thin" is the characterization of action that is involved in agentive awareness (see Bayne & Pacherie, 2007). At the thin end of the spectrum, the action description is minimal, asserting merely that the agent is acting, as opposed to moving passively. At the thick end of the spectrum, the description is rich, and might involve details about what the agent is doing, e.g., drinking coffee, and how, e.g., reaching and grasping for her mug. A theory of agentive awareness should be able to account for this representational character, and perhaps even specify whether it is thick or thin.

In sum, the five aspects of agentive awareness that must be explained by any theory are its self-referential character, its subjectively unmediated character, its phenomenal character, its representational character, and what happens when it breaks down in disorders of volition. Next, I outline three questions that will help us approach the project before us.

§3 Three Questions To Frame Theorizing

Theorizing about agentive awareness revolves around three main questions, which will also individually serve as the focal points of Chapters 2, 3, and 4 of this dissertation, with Chapter 5 bringing together the insights arrived at in each of those chapters into one positive account. The three questions are as follows: (i) What type of awareness is

agentive awareness? (ii) How does agentive awareness relate to action control? (iii) What is the psychological mechanism that gives rise to agentive awareness?

These questions heavily intertwine. An answer to any one of them restricts or even determines the answers to the other two, so they cannot be treated completely in isolation. But each one serves to highlight a feature of agentive awareness that has been hotly debated in the literature. I expand below.

§3.1 What Type of Awareness Is Agentive Awareness?

In general, an individual is aware of something in virtue of being in suitable mental states. One is aware of the greenness of a lime, for example, in virtue of being in an appropriate visual sensory state. And one is aware that Beethoven's *Moonlight Sonata* is playing in virtue of being in an appropriate auditory perceptual state. These are examples of mental states with sensory qualities providing one with awareness of physical properties or states of affairs. In addition, one may also arguably be aware of something in virtue of having a suitable thought about it (see Rosenthal, 2004). In particular, a thought about something as being present may serve to make one aware of that thing, even if one is not at that moment in perceptual contact with it.

Since agentive awareness is a form of awareness, namely awareness *of* oneself as acting, we are confronted with the question of what kind of mental state or states are responsible for such awareness. In general, I will adopt a disjunctive characterization of mental states, according to which a mental state is a state that has *intentional properties* (e.g., beliefs, desires, intentions) or *sensory properties* (e.g., perceptions, sensations) or

motoric properties (e.g., motor commands).[5] A state may have more than one of these types of property, e.g., perceptions have both intentional properties and sensory properties, but whatever the case about that, a state is not considered to be mental unless it has at least one of these types of property.

There are, of course, many important and unresolved questions as to how to understand these properties—intentional, sensory, and motoric. It is not my project to settle them, or even engage with them here, but rather to adopt a framework within which I can work out a theory of agentive awareness. As such, I will endorse without argument the view that a mental state is *intentional* just in case it has *mental attitude* (e.g., belief, desire) directed towards *propositional content* (e.g., that it is raining). And I will also take on board the view that sensory qualities of mental states have two main features: (i) they belong to representational states that are the outputs of sensory modalities, and (ii) they correspond to sensory properties of objects in the world (e.g., color, shape, texture), such that when a healthy creature senses these properties under normal conditions, the relevant sensory modality outputs a state with the relevant sensory quality or qualities. I will call mental states with sensory properties *sensory states*. Finally, I will view motoric properties of mental states as having two main features: (i) they belong to representational states that are the outputs of motor systems, and (ii) they correspond to motoric properties of bodily movements (e.g., direction, velocity, amplitude). I will call mental states with motoric properties *motoric states*.

[5] Motoric states are not traditionally included in the taxonomy of the mental, but this is arguably due in large part because they fail to be conscious, and so escape the notice of many theorists. Still, there is good reason to include them as a category of the mental, as they play a key role in our understanding of action production (see Pacherie, 2011).

With this rough taxonomy in hand, we may identify the main options for the types of mental state in virtue of which an individual is agentively aware. There are four such options: (i) cognitive states (e.g., beliefs, thoughts, predictions), (ii) conative states (e.g., intentions), (iii) sensory states (e.g., sensations, perceptions, sensory images, emotions), and (iv) motoric states (e.g., motor commands).

In the course of the dissertation, I will consider each of these options, ultimately defending the view that agentive awareness is a type of cognitive awareness—in other words, when one is agentively aware in virtue of being in suitable cognitive states. In Chapter 2, I will argue that the sensory approach to agentive awareness is unworkable. It will emerge from my discussion in Chapter 3 that motoric states are also inappropriate candidates for providing one with agentive awareness. In Chapter 5, I will evaluate both the view that agentive awareness is a type of conative awareness—specifically, grounded in intentions to act here and now—and the view that it is a type of cognitive awareness that is *based* on such intentions, settling on the latter.

§3.2 How Is Agentive Awareness Related to Action Control?

Another theoretical angle from which agentive awareness is commonly approached concerns its relation to action control. Before fleshing this out, I will make some brief remarks on the model of action that I will adopt throughout the dissertation, which is also intimately tied to control—what I call the *control theory of action*. Central to such a theory is the claim that the relationship of control that holds between an agent and her bodily movements as they unfold is what grounds the distinction between action and non-action, i.e., between what an agent does, and what merely happens to her. In its simplest formulation, the main claim of the theory is that a bodily movement is an action

if and only if it is under the agent's control.

The control theory of action may be viewed as an expansion of the traditional *causal theory of action*, which looks only to psychological causal *antecedents* of bodily movement, such as reasons or intentions, for the distinguishing features of action (e.g., Davidson, 1963/2001a). On a causal theory of action, what happens *at the time during which* bodily movements are being carried out is of no relevance to the question of whether they are actions; that is to say, the question has been settled before the bodily movements have begun.

Given the dynamic nature of action, the focus of the causal theory on causal antecedents of action seems too narrow—it leaves out an important feature of actions as events that unfold in time. A control theory of action seeks to remedy this by expanding the conditions on something's being an action to include the causal interactions characteristic of action control (see Frankfurt, 1978).

Generally speaking, action control is composed of two main interacting functions: (i) guidance, and (ii) monitoring. An agent *guides* her bodily movements if and only if (a) her goal states (intentions or motor commands) suitably cause her body to move in the ways that they specify, and (b) she is disposed to compensate for execution errors registered on the basis of monitoring states. In turn, an agent *monitors* her bodily movements if and only if she is in monitoring states (sensory states like sensations or perceptions) that accurately represent her bodily states.

Action control is commonly characterized as a hierarchical process, with each level and stage being associated with different mental states and processes, working together to successfully guide and monitor action (e.g., Jeannerod, 1997; Pacherie, 2008).

There are two main levels of action control: *intentional control* and *sensorimotor control*. Intentional control concerns deliberation, planning, and anchoring the action plan in the agent's present context. At the level of intentional control, an agent conceptualizes her actions and sensory environment by way of intentions and perceptual states. Sensorimotor control by contrast, concerns the detailed execution of action in the face of kinematic and sensory processing constraints. At this level of control, conceptualization is not operative—the agent implements the guidance and monitoring of her movements by way of motoric and sensory states that do not carry conceptual content.[67] I will view the distinction between actions and non-actions as that between bodily movements that are under an agent's sensorimotor control, and those that are not. And I will view the distinction between intentional and non-intentional actions as the distinction between those bodily movements that are under an agent's intentional control, and those that are not.

---

[6] See Pacherie's (2006, 2008) *dynamic theory of intentions*, which offers a rich theoretical framework within which to understand the different levels of action control that I am sketching here. My own picture is greatly influenced by Pacherie's elegant treatment of action control.

[7] In positing both a higher, intentional level of control, and a lower, sensorimotor level of control, I am endorsing what is known by control theorists as a *dual control model* (e.g., Humphreys & Riddoch, 2003; Perner, 2003). Perhaps the most widely-discussed version of a dual control model is that developed by Norman and Shallice (1986) who distinguish between two systems implicated in action control: the supervisory attentional system (SAS) and the contention scheduling system (CSS). The latter stores "response schemas", which are sequences of stimulus-response pairings that are partly activated by environmental triggers, and partly activated by the SAS. The role of the SAS is to perform "executive functions" of action planning, monitoring, and inhibition of habitual responses where inappropriate. Norman and Shallice leave the SAS to function largely as a homunculus, however, without taking care to specify how it implemented its various functions. Still, the general structure of this model has been largely influential in more recent discussions of action control, including the one presented here.

A commonsense example will help to further illustrate how the two types of control may work in tandem to execute actions. Suppose you are following a recipe for a pasta dinner. At the level of intentional control, you will need, among other things, to plan to follow the order of steps in the recipe, to identify by way of perception the correct ingredients in your kitchen, to determine the appropriate boiling time for the pasta in accordance with the pasta type and your aesthetic preferences, and so on. But in order to implement each of these steps, you will also need to execute the appropriate bodily movements—and this is where sensorimotor control comes in. Sensorimotor control ensures, for example, that you form a grip aperture that is a suitable size for grasping the handles of the pasta pot, that you move your arm with the correct trajectory to reach for the salt, and that you shift your body weight and posture so as to keep your balance.

We can further get a grip on the distinction between intentional control and sensorimotor control by identifying some general markers of each. Intentional control is relatively slow, flexible, and typically involves conscious states and processes. Sensorimotor control, on the other hand, is relatively fast, rigid, and typically involves non-conscious states and processes.

There is also evidence that intentional control and sensorimotor control have unique neuroanatomical profiles, operating by way of distinct brain circuits and regions. It is thought that these circuits converge on the primary motor cortex (M1), which sends motor commands to the spinal cord and muscles (Haggard, 2008). One source of input to M1 comes from the SMA, which itself receives inputs from the basal ganglia and prefrontal cortex, and which is thought to play a key role in implementing the flexibility that is a feature of intentional control (see Jeannerod, 1997, p. 149). Some evidence for

this comes from clinical cases. For example, recall that those with utilization behavior, which is thought to result from damage to the SMA, show repetitive behavior that is heavily responsive to external stimuli—they show a lack of inhibition, and a lack of flexibility in their responses, which are both features of intentional control. In addition, neuroimaging studies investigating the neural basis of "self-paced" actions, that is, those that an individual decides when to perform, versus "stimulus-driven" actions, that is, those that are dictated by the occurrence of a stimulus, show significantly stronger activation in the pre-SMA for the former type of action over the latter (Deiber, Honda, Ibanez, Sadato, & Hallett, 1999; Jenkins, Jahanshahi, Jueptner, Passingham, & Brooks, 2000). Other neuroimaging studies suggest that the pre-SMA region is responsible for the inhibition of ongoing tasks, and for selecting among competing tasks and switching between them (e.g., Nachev, Wydell, O'Neill, Husain, & Kennard, 2007).

Another key input to M1 comes from the pre-motor area of the brain, which itself receives inputs from parietal regions (Haggard, 2008). This area seems to be more active in stimulus-driven responses, and in the visual guidance of object-oriented responses, and is therefore more likely responsible for sensorimotor control (Chen, Thaler, Nixon, Stern, & Passingham, 1995).

All this emphasis on the ways in which intentional control and sensorimotor control differ should not give the impression that they operate separately—they typically work together to successfully guide our actions to completion. This is because, as Jeannerod (1997) notes, the simple actions that are *exclusively* in the realm of sensorimotor control are rarely performed for their own sake except in laboratory settings. They are typically embedded in more complex action plans that are determined

at the level of intentional control. For example, one's basic action of reaching and grasping an object may be one step in the larger action plan of passing the salt to one's dinner companion. In this way, sensorimotor control is frequently shaped by intentional control.

A simple experiment will help to bring this point home. Marteniuk, MacKenzie, Jeannerod, Athenes, & Dugas (1987) had participants reach and grasp a small disc, which, in one condition ("fit" condition) they had to place into a tight spot, and in another condition ("throw" condition) they had to throw into a large container. The kinematic properties of the reach were different depending on what the overall intention was—whether it was to fit or to throw the object. But, as Jeannerod (1997) observes in his discussion of this study, "[t]hese changes are not predicted by the kinematic requirements, nor by the coordination constraints of the movement itself" (126). Rather, they are determined by the higher-level goals of the agent, as specified at the level of intentional control.

This is far from a complete treatment of either sensorimotor control or intentional control. Nor is it even close to a full discussion of the nature of action. Still, I hope the foregoing is sufficient to offer a working model of control and action for the remainder of this dissertation.

Given this background, then, the question before us with respect to agentive awareness is, as Gallagher (2007) puts it, "[s]hould we think of the sense of agency as belonging to the realm of motor control and body movement, or as belonging to the realm of the intentionality of intentional action?" (p. 349). On what I will call *low-level accounts*, agentive awareness is viewed as a function of the states and processes involved

in sensorimotor control. On *high-level accounts*, by contrast, it is a viewed as a function of the states and processes involved in intentional control. In addition, there are accounts that adopt a mixed approach, taking agentive awareness to arise out of both high-level and low-level states and processes, which I will call *integrated accounts*.

In Chapter 3, I argue against the claim that the states and processes involved in low-level, sensorimotor control contribute to agentive awareness, thus making an indirect case for high-level accounts, which take the states and processes involved in intentional control, pitched as it is in terms of conceptual states like intentions, thoughts, and perceptions, as their starting point. Here, I wish to lay down the foundation of the low-level approach, since it is the backbone of much theorizing in agentive awareness and will play a role throughout this dissertation.

By far, the most widely adopted theoretical model of sensorimotor control, and the one that is most frequently appealed to in order to explain agentive awareness by low-level approaches, is the comparator model (CM) (e.g., Bayne & Pacherie, 2007; Carruthers, 2012; de Vignemont & Fourneret, 2004; Frith, 1992; Gallagher, 2000; Jeannerod, 2006; Pacherie, 2008; Synofzik, Vosgerau, & Newen, 2008). The CM in its present form has emerged in part from a long-standing controversy among theorists about the nature of sensorimotor control. On either side of the debate are two competing models: (i) *feedforward models*, and (ii) *feedback models*. These models differ based on how much weight they assign to the respective roles of motor commands and sensory feedback in the guidance of movement.

Feedforward models appeal primarily to motor commands in order to explain how we guide our actions. On such models, when I reach towards an object to grasp it, for

example, the motor system issues motor commands prior to the onset of movement, which mostly specify its various stages. Sensory feedback is used in the guidance of movement only towards the end, near completion of the movement, when its velocity has decreased. On feedback models, by contrast, the stages of the movement are not programmed prior to its initiation; motor commands are constantly updated and adjusted by way of sensory feedback as the movement unfolds.

Against feedback models, opponents argue that there is a delay in sensorimotor loops that makes such models untenable (Wolpert, Ghahramani, & Jordan, 1995). More specifically, they hold that the minimum delay required for visual or proprioceptive information to influence an ongoing movement is too long to be used for the guidance of movement. As support for this claim, appeals are made to estimates on the basis of behavioral studies, which suggest that the minimum delay for visual or proprioceptive feedback is 80 – 100 ms, whereas the average duration of a visually guided reaching movement is 300 – 700 ms (Desmurget & Grafton, 2000; see also Wolpert & Miall, 1996 for slightly varying estimates). This suggests that such a movement would be largely underway before any visual or proprioceptive feedback could begin to influence it. Generalizing to all movements, theorists have taken this as evidence that, though sensory feedback may play some role, it is not playing the central role in sensorimotor control suggested by feedback models (Desmurget & Grafton, 2000).

Forward modeling is thought by some to offer an intermediary approach between both feedforward models and feedback models. The main function of forward modeling is to form predictions of the consequences of motor commands issued by the motor system. Theorizing about forward modeling has been inspired by the independent work

of the 19th century psychologists Bell (1868/1974), Purkinje (1825), and von Helmholtz (1866), who each sought to understand how it is that the world appears stable in visual perception despite constant movement from the retina. They noted that when one presses one's eye, the world appears to move, while when one moves one's eye voluntarily, the world appears to remain stable. To explain this puzzling observation, they theorized that when one voluntarily moves one's eyes, the brain produces a copy of the motor signal sent to the eyes, and uses it to compensate for the retinal image motion that follows by canceling it out. Since no such copy is generated in the case of passive eye movement, no canceling out occurs, and the world itself appears to move (Bridgeman, 2007).

In the 1950s, this early work was extended, and the posited copy of the motor signal was given a name; von Holst and Mittelstaedt (1950) dubbed it the 'efference copy' (Bridgeman, 2007). Since then, the idea of efference copy has been generalized from eye movements to all sensorimotor control, and used to inform the debate between feedforward and feedback models. Theorists have suggested that the forward model takes as input the efference copy, and based on this and an internal model of the motor system and environment, it outputs a prediction of the sensory and motoric consequences of motor commands (Wolpert, 1997).[8] These forward model predictions are thought by some to help explain how it is that we are able to exert online control over rapid, fine-grained movements despite delays in sensory feedback. Rather than wait for the actual

[8] This predicted sensory feedback is sometimes called 'corollary discharge' (Sperry, 1950).

sensory feedback, the motor system employs the predictions outputted by the forward model (Wolpert & Miall, 1996).[9]

The CM is largely grounded in this theoretical framework for sensorimotor control (see Figure 1). When specifying its components, theorists include both a forward model and what is known as an inverse model. The role of the inverse model is to compute appropriate motor commands for driving the body from its current state to a given goal state. It does this by integrating sensory information about the current state of the agent and the environment. The forward model, as sketched above, works in the opposite direction, taking the efference copy—itself a duplicate of the motor command outputted by the inverse model—as input, and outputting a prediction of the sensory consequences that the movement specified by the efference copy would result in.

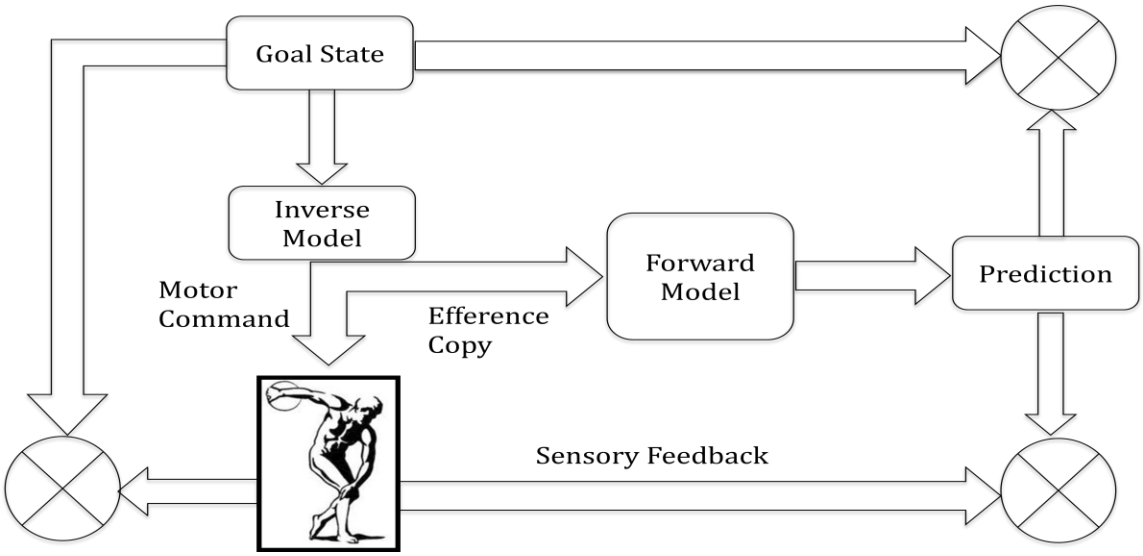

Figure 1. The Comparator Model.

In addition to this, three separate comparisons involving the goal state of the agent, the forward model prediction, and the resulting sensory feedback from the

[9] The main site of forward models in the brain is thought by some to be the cerebellum (Blakemore, 2002; Blakemore, Wolpert, & Frith, 1998; Blakemore, Wolpert, & Frith, 1999).

movement are posited in order to explain how it is that we successfully control our movements.

The first such comparison is the most widely recognized by theorists, and is between the forward model prediction and the actual sensory feedback resulting from the movement (Blakemore & Frith, 2003; Blakemore, Frith, & Wolpert, 1999; Blakemore, Oakley, & Frith, 2003; Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Blakemore, Wolpert, & Frith, 2000; Blakemore, Wolpert, & Frith, 2002). If the comparison between what is predicted and what actually happens results in a mismatch, an error signal is sent to the forward model to help train it up for future movements.

A second comparison is posited between the prediction outputted by the forward model and the goal state (Blakemore, Wolpert, & Frith, 2002; Desmurget & Grafton, 2000). This comparison is thought to take place prior to the onset of bodily movement. If there is a mismatch here, then an error signal is sent to the inverse model and a new motor command is programmed. This is thought to enable the correction of a movement before feedback from the movement is actually available, which is an advantage in the face of aforementioned delays inherent in sensorimotor loops. It is also thought to help train up the inverse model, to program more accurate motor commands.

Finally, a third comparison is posited between the goal state and the sensory feedback from the movement. If a mismatch occurs here, then the inverse model is updated in order to improve its accuracy. This comparison is thought to be important for motor learning, since it allows the system to register errors in action execution, and thereby take steps to enhance future performance (Pacherie, 2008).

As I mentioned, many theorists have found the view that agentive awareness is somehow grounded in the states and processes of the CM attractive. When it comes to explaining agentive awareness, it is most frequently the comparison between the forward model prediction and the sensory feedback from the movement that is thought to do the work (e.g., Bayne & Pacherie, 2007; Pacherie, 2008; Bayne, 2011b). As Bayne (2011b) puts it, "when there is a match between predicted and actual state, the comparator sends a signal to the effect that the sensory changes are self-generated; when there is no match (or an insufficiently robust match), sensory changes are coded as externally caused" (495).

In light of this background, my goal in Chapter 3, more specifically, is to argue that the CM approach to agentive awareness, which takes the above comparison as central, does not succeed. My case against this popular view has two parts: First, I argue that much of the evidence adduced in favor of the approach does not, upon closer examination, hold up. Second, I argue that there are insurmountable difficulties for any approach, like the CM approach, that takes agentive awareness to be grounded in sensorimotor control, since this level of control frequently dissociates from agentive awareness. This paves the way for the high-level account of agentive awareness I go on to develop in Chapter 5 in terms of intentions and thoughts.

§3.3 What Is the Psychological Mechanism Underlying Agentive Awareness?

Another question that guides research on agentive awareness concerns *how* exactly agentive states come about at the psychological level. The dominant view is that agentive awareness arises out of some kind of matching process, wherein goal states are compared with action-monitoring states, and if they match in content, then agentive

awareness arises. I just introduced a low-level version of this view when discussing the CM approach to agentive awareness—on that variant, the goal states are forward model predictions and the monitoring states are sensory states corresponding to the bodily movement and its consequences. But high-level matching accounts have been put forward as well, perhaps most famously by Wegner (2002). On Wegner's view, intentions are compared with perceptions of action, and if there is a match—along with some other conditions being met—one infers that one's intention causes one's action, and one subsequently has agentive awareness over one's action.[10]

In Chapter 4, I challenge Wegner's high-level matching account of agentive awareness. First, I will evaluate the empirical evidence cited in its favor and conclude that it is wanting. Next, I will argue that agentive awareness often arises in the absence of any match between goal states and monitoring states, so such a match is not required. Indeed, a problem with Wegner's view is that it attributes to agents a highly unnatural stance towards their own actions, on which they find out about them in the same way that they find out that they are the causes of effects in the external world. I argue that this is the wrong way to view agentive awareness.

§4 Conclusion

My arguments in Chapters 2, 3, and 4, framed in terms of the questions I have laid out in this chapter, together lead to the doorstep of a theory of agentive awareness that

[10] Oftentimes, the CM matching model is referred to as a *predictive model* and the Wegner-style matching model is referred to as a *retrospective model* (e.g., Chambon, Wenke, Fleming, Prinz, & Haggard, 2013; Haggard, 2005; Pacherie, 2007; Wenke, Fleming, & Haggard, 2010). I find these labels misleading, however, since both approaches involve predictive (i.e., intention or forward model predictions) and retrospective (i.e., an inference that the pre-action and action state match, or a low-level registration that they match) components.

takes it to be a form of non-sensory awareness, involving the states operative in intentional control, and requiring no match between such states and any sensory feedback. In particular, I will defend a view on which agentive awareness is a type of cognitive awareness. What it is to be agentively aware of oneself as A-ing is to have a suitable thought to the effect that one is A-ing. Or so I will argue.

Before moving on, I register some caveats. First, I will focus exclusively on unitary accounts of agentive awareness, that is, on accounts that take such awareness to be one type of phenomenon. This starting assumption is at odds with some of the recently emerging views on agentive awareness that take it to be a multifaceted phenomenon. For example, Synofzik et al. (2008) defend what they call the *multifactorial weighting model* of agentive awareness, on which:

> Action-related perceptual and motor information—such as efference copies, sensory feedback modalities, and their comparison—is rather entered into a multifactorial weighting process bringing together different sense modalities (e.g. intermodal congruencies), efferent-afferent congruencies, and different levels of cognitive processing that indicate agency of actions for a person... (p. 226)

On this type of view (see also Pacherie, 2008), various action-related cues get weighted within a Bayesian framework to yield agentive awareness, so the unitary accounts that I evaluate in this dissertation would be viewed as not being mutually exclusive, but rather as all getting at an aspect of the processes involved in generating agentive awareness. In a similar spirit, Pacherie (forthcoming) argues that once one appreciates the dynamic character of agentive awareness, that is, as a type of awareness that unfolds in time as an action is executed, there is reason to view agentive awareness as a heterogeneous

phenomenon that may shift in its intentional structure as well as its level of grain, so that it is captured by non-sensory and sensory states, high-level and low-level states alike.

While such models may indeed be promising, and serve to nicely integrate the growing literature on agentive awareness and its many complexities, it is worth exploring all credible options before abandoning a unitary view of agentive awareness. And the type of unitary account I will later defend not only is not yet represented in the literature, but, I will argue, can accommodate a wide range of theoretical data. If, in the fullness of time, a multifaceted view of agentive awareness ends up being a stronger approach, at the very least the account I develop here will help to illuminate one of those many facets.

Second, it is worth making a brief remark on consciousness, since it is of course related to agentive awareness, given that it is a type of self-consciousness. When it comes to competing theories of consciousness, I will remain neutral in the dissertation. The account I develop of agentive awareness is meant to be compatible with all the main alternatives in the literature, including higher-order theories (e.g., Lycan, 1996; Rosenthal, 2005), attention-based theories (e.g., Prinz, 2012), global workspace theories (e.g., Baars, 1988; Dehaene & Changeux, 2011), and first-order theories (e.g., Block, 2007; Dretske, 1995). Whether one is agentively aware or not, on my view, will depend on whether one is in the appropriate mental state. And one should be able to plug in one's favorite theory about what makes a mental state conscious without any relevant difference to the view that I develop.

CHAPTER 2: THE SENSORY APPROACH TO AGENTIVE AWARENESS

§1 Introduction

A promising way to get a grip on the nature of agentive awareness is to ask what *kind* of awareness it is. As I remarked in Chapter 1, in general, one is made aware of things by being in appropriate mental states. One's sensory states, for example, make one aware of one's body and objects in one's physical environment. And, similarly, one can sometimes be aware of something by being in suitable cognitive states, such as having an occurrent thought to the effect that something is present. But where in our mental architecture—cognition, perception, or elsewhere—does *agentive* awareness lie?

It is plain that one sometimes has thoughts about one's own actions. Perhaps most commonly, one has thoughts about *how* one is acting, for example, how one is flipping a pancake, or drawing a picture. But one sometimes also simply thinks *that* one is performing a certain action—for example, that one is making pancakes, or that one is drawing. I will call such thoughts *agentive thoughts*.

Though many acknowledge that we sometimes have agentive thoughts, rarely do theorists defend the view that agentive awareness is, at bottom, a matter of having such thoughts. Rather, it is commonly held that agentive thoughts are based on some *antecedent* awareness of oneself as acting. And the most frequently endorsed view is that such antecedent awareness is best understood as a form of *experience*. For example, Bayne and Pacherie (2007) write:

> We hold that the 'vehicles' of agentive self-awareness are often more primitive than judgments. Think of what it is like to push a door open. One might judge that

> one is the agent of this action, but this judgment is not the only way in which one's own agency is manifested to oneself; indeed, it is arguably not even the primary way in which one's own agency is manifested to oneself. Instead, one *experiences* oneself as the agent of this action. (p. 476)

And, in a similar vein, Bermudez (2010) writes:

> [B]eliefs about our own agency need to be anchored in something. They are not free-standing and they do not come out of nowhere. Beliefs are formed for reasons and the most plausible candidates for those reasons relate to our experiences as agents. (p. 592)

So a common way to model the psychology surrounding agentive awareness is to suppose that agentive thoughts are based on agentive experiences. But this just pushes the question back a step, for now we must ask: what kind of experiences are agentive experiences? After all, the term 'experience' is often used simply to describe a mental state with some phenomenal character, and this could apply to a wide variety of mental state *types*, including sensations, perceptions, emotions—and even, on some views, thoughts (Pitt, 2004). Given this liberal application of the term, claiming that agentive awareness bottoms out in agentive experiences does not help us arrive at an answer to the question with which we began of what *kind* of awareness it is.

A more substantive proposal has recently been defended by Bayne (2011b), according to whom agentive experiences are best understood as *sensory* experiences. As Bayne (2011b) puts his own version of the view:

> Just as we have sensory systems that function to inform us about the distribution of objects in our immediate environment, damage to our limbs, and our need for

food, so, too, we have a sensory system (or systems), whose function it is to inform us about facets of our own agency. (p. 491)

In this chapter, I will evaluate the sensory approach to agentive awareness. Of course, a proponent of the view that agentive experiences are sensory experiences need not follow Bayne in supposing that they arise out of a dedicated, *sui generis* sensory modality. In addition to Bayne's position, I will consider another popular version of the sensory approach that does not take this commitment on board.

The chapter is organized as follows: In section 2, I lay out the characteristics that agentive experiences must have, at a minimum, if they are to be understood as sensory experiences—what I call *sensory agentive experiences* (SAEs). In section 3, I critique two of the main arguments that have been offered in favor of positing SAEs. In section 4, I advance my positive argument against the view that agentive experiences are sensory experiences. The argument runs as follows: If agentive experiences are sensory experiences, then they must exist in some sensory modality (or set of sensory modalities). This sensory modality (or set of sensory modalities) must either be one with which we are familiar (e.g., vision, audition, olfaction, gustation, proprioception) or a novel sensory modality (or set of sensory modalities). I argue that neither option is workable. In the absence of any modality within which to locate SAEs, I conclude that there are none, and that we must look to non-sensory models of agentive awareness.

§2 What Are Sensory Agentive Experiences (SAEs)?

To be clear, in denying that there are SAEs, I am not denying that we sense our actions. We regularly sense our bodily movements. And actions are often, if not always, bodily movements. Nor am I denying that engaging in action involves rich sensory

experiences. When we act we often have robust visual experiences of our bodies in motion, and proprioceptive experiences of our limb positions and movements, as well as the activity of our muscles and joints. In addition, acting sometimes *shapes* our overall sensory experience in more subtle ways. There is evidence, for example, that events following our actions appear subjectively to occur closer in time than those following our non-actions, even if the temporal interval is, in fact, the same—a phenomenon known as the *intentional binding effect* (Haggard, Clark, & Kalogeras, 2002).[11] I am also not contesting here that there is a *phenomenology* of agency. Even if there are no SAEs, this leaves open the possibility that acting has a distinctive phenomenology that arises from being in the relevant non-sensory states.

What I do dispute, however, is that we come to be aware of ourselves as acting by way of any sensory experience or set of sensory experiences. As tempting as it may be to assimilate agentive awareness to a form of sensory awareness, we should resist doing so. Before making the case for this claim, however, it will be useful to get a better handle on what it means for something to be a sensory agentive experience in the first place.

Broadly speaking, agentive experiences, whether sensory or not, are thought to have two main features. First, in line with the profile of agentive awareness more generally, they are widely viewed as states that *represent* the way the world is—they are "richly intentional" (Horgan, Tienson, & Graham, 2003, p. 333; see also Bayne, 2011b; Bayne & Pacherie, 2007; Carruthers, 2012; Pacherie, 2008; Synofzik, Vosgerau, & Newen, 2008). More specifically, they represent *oneself* as being a certain way—as being the source of a present action. They are not, therefore, to be construed as purely "raw

[11] I will discuss this phenomenon in further detail in Chapter 3.

feels" without representational properties, the way some theorists, both traditional and contemporary, understand qualia, for example (e.g., Block, 2003; Reid, 1764/1997).

The second feature is one that I have already gestured towards in my earlier remarks: agentive experiences, if they exist, are the basis for agentive thoughts (Bayne & Pacherie, 2007; Bermúdez, 2010; Synofzik et al., 2008). When one has a thought to the effect that one is acting, on this view, it is in some way derived from or dependent upon an antecedent experience of oneself as acting. The nature of the relationship between experiences in general and corresponding thoughts or judgments, especially its justificatory status, is a hotly debated topic (for the relevant debates see, e.g., Armstrong, 1968; McDowell, 1994; Peacocke, 2001). I will remain neutral on the details of how it works here. Whatever the case about that, it bears emphasizing that given their putative role in serving as the basis for agentive thoughts, agentive experiences are most reasonably construed as the way by which we *come to be* aware of ourselves as acting. They are not *merely* states that provide such awareness—agentive thoughts satisfy that condition as well—they are the *first stop* for agentive awareness, the *primary* way by which we become so aware.

This last point is worth highlighting for the reason that, even if it were to turn out that there are sensory states that represent oneself as acting, this would not suffice to show that agentive awareness is sensory awareness. For these sensory states might themselves derive from some antecedent agentive awareness that is non-sensory in form. And if so, then agentive awareness is, at bottom, a non-sensory form of awareness. Analogously, even though we sometimes form beliefs about the colors of certain objects, e.g., the belief that there is red, it would be a mistake to construe color awareness as a

form of purely cognitive awareness. Rather, since these perceptual beliefs are based in some way on red and blue visual sensations, the right thing to say is that color awareness is, at bottom, a form of visual awareness. Similarly, it is the states by which we first come to be agentively aware that are of primary concern, not states further downstream in our psychological processing that depend on them.

So far I have listed two general features that theorists typically attribute to agentive experiences, whether or not they are sensory. But if agentive experiences are sensory experiences in particular, then they must also possess sensory qualities. Sensory qualities are presumably the features that distinguish SAEs, if they do indeed exist, from the agentive thoughts that they are supposed to support. The literature on sensory qualities is extensive and knotty. I will therefore attempt to stay as neutral as possible on the nature of sensory qualities—the arguments I present in this chapter will not depend on adopting any particular account. To that end, as I mentioned in Chapter 1, I will characterize sensory qualities as having two main features, which I take to be uncontroversial: (i) the states to which they attach are the outputs of sensory modalities, and (ii) they correspond to sensory properties of the body or objects in the world, such that when a healthy creature senses these properties under normal conditions, the relevant sensory modality outputs a state with the relevant sensory quality.

Some hold that the sensory qualities of experiences are to be identified with the representational properties of the experience (e.g., Harman, 1990; Tye, 2000). Others maintain that sensory qualities are something over and above the representational properties of experience (e.g., Block, 2003; Chalmers, 2004). Still others prefer to identify them with represented properties of *objects* or, at one extreme, to deny their

existence altogether (e.g., Harman, 1990). Once again, I will not take a stance on these issues, except to hold that sensory qualities do indeed exist. If agentive awareness is sensory awareness, we should be able to identify some type of sensory quality or set of sensory qualities, as characterized above, uniquely associated with it. A type of sensory quality that sometimes accompanies awareness of oneself as acting and at other times accompanies awareness of oneself as passively moving will not do here, as this cannot then serve to aid in the discrimination of actions and passive movements, which is required if agentive experiences are to be the basis for agentive thoughts.

The goal of this chapter is to argue that SAEs, understood as states with these three features, do not exist. There is no type of mental state that represents oneself as acting, possesses sensory qualities, and is the primary way by which we come to be aware of ourselves as acting, thereby serving as the basis for agentive thoughts. Of course, not all are convinced that this is the case. I turn now to consider two prominent arguments in favor of positing SAEs.

§3 The Arguments From Pathology and Cognitive Impenetrability

It might strike one as odd to require *arguments* to establish the existence of a certain kind of sensory experience. After all, one might reason, experiences are directly introspectible, and if so, what good are arguments? As Searle (1983) remarks:

> It is a bit difficult to know how one would argue for the existence of perceptual experiences to someone who denied their existence. It would be a bit like arguing for the existence of pains: if their existence is not obvious already, no philosophical argument could convince one. (p. 44)

When it comes to agentive experiences, however, and as I emphasized in Chapter 1, many have supposed that they are *not* readily available to introspection, at least to the same degree as familiar sensory experiences like color experiences. Recall that they are, instead, frequently described as "thin" and "evasive" (Metzinger, 2006).

Some may embrace a skeptical conclusion on this basis, holding that the purported character of agentive experiences should lead us to doubt that they are sensory experiences in particular. After all, what other type of sensory experience can be considered "thin" and "evasive" in this way? Certainly not pains, colors, smells, sounds, and the large variety of robust sensory experiences with which we are acquainted. The striking contrast between rich sensory experiences and impoverished agentive experiences may lead one to worry that they are not, in fact, akin.

But this is not sound reasoning. First of all, the fact that a member of a class does not share some feature or set of features to the same degree as other members of that class is not sufficient reason, of course, to deny its membership in that class. Plainly, objects with faded colors are still colored objects, and very faint sounds are still sounds. Second, even if we were to lack *complete* introspective access to agentive experiences, which is not the claim being entertained here, this would not entail that they do not exist. This would follow only if introspection gives us exhaustive access to our sensory states, and we have good independent reason to reject this Cartesian conviction from cases of subliminal perception and blindsight (Stoerig & Cowey, 1997; see also Rosenthal, 2010 for discussion). Moreover, we have other ways of becoming aware of our experiences that do not rely on directly introspecting them. For example, we can establish their existence inferentially (Tye, 2003). I might infer from my ability to sort color chips

according to their different hues that I am experiencing the hues of those chips. And I could do this even if I could not directly introspect any features of those experiences. So introspection neither confirms nor disconfirms the existence of SAEs. We must look instead to theory. In the remainder of this section, I consider two arguments—the first in favor of positing agentive experiences more generally, and the second in favor of positing SAEs in particular.

§3.1 The Argument from Pathology

A popular line maintains that sound explanations for certain pathologies of agency make an indispensible appeal to agentive experiences, and that this gives us good reason to posit them.[12] Recall anarchic hand syndrome, in which individuals execute complex, goal-oriented movements with an upper limb, usually their hand, all the while denying authorship of those movements (Marchetti & Della Sala, 1998). The neurologist Sergio Della Sala (2005) describes a particularly memorable outing with one of his anarchic hand patients:

> One evening we took our patient, Mrs GP, to dinner with her family. We were discussing the implication of her medical condition for her and her relatives, when, out of the blue and much to her dismay, her left hand took some leftover fish-bones and put them into her mouth. A little later, while she was begging it not to embarrass her any more, her mischievous hand grabbed the ice-cream that her brother was licking. Her right hand immediately intervened to put things in

[12] Though this first argument, if it were successful, would not by itself be sufficient for establishing the existence of SAEs, since the relevant experiences could be non-sensory in nature, it is worth considering here, as some theorists may be tempted to appeal to it as a stepping stone on the way to arguing for SAEs.

> place and as a result of the fighting the dessert dropped on the floor. She apologised profusely for this behaviour that she attributed to her hand's disobedience. Indeed she claimed that her hand had a mind of its own and often did whatever 'pleased it'. (p. 606)

Mrs. GP does not identify as the agent of the movements of her anarchic limb, no matter how much they appear to be her actions from the outside. She is aware of them passively, and does not take them to express her own agency.

Some theorists regard anarchic hand syndrome, and other cases in which one feels alienated from one's actions, as providing evidence for agentive experiences that normally accompany our actions. For example, concerning such cases, Bayne (2011b) writes:

> It seems plausible to appeal to agentive experience—or the lack thereof—in order to *explain* why [anarchic hand patients] deny having performed the anarchic actions. Surely it is the fact that the normal and expected experience of *doing* has been replaced by an experience of *happening* that leads these patients to judge that the action is not theirs. (p. 498, emphasis in original)

Bayne (2011b) is here urging that a reasonable explanation for why anarchic hand patients like Mrs. GP deny ownership of their anarchic movements is that they have what we might call *experiences of passivity* replacing their agentive experiences, and driving these reports.

This diagnosis is problematic for several reasons, however. First, the existence of experiences of passivity does not entail that there are agentive experiences. This would be like inferring on the basis of experiences of pain that whenever one is not having such

experiences, one is having experiences of painlessness. It may be that whenever something goes wrong with our own actions, we come to have experiences of passivity, but that when our actions are carried out correctly, no corresponding agentive experience is present.

Second, it is worth stressing that even if the existence of experiences of passivity were to entail the existence of agentive experiences, it would not follow that these experiences are both sensory in form. In the same way that it is problematic to infer the existence of agentive experiences from the existence of experiences of passivity, it is problematic to draw inferences concerning the *properties* of agentive experiences from claims concerning the *properties* of experiences of passivity. Even if experiences of passivity are sensory, it may be that agentive experiences are not—they could instead be purely cognitive or affective experiences, for example. We would require independent reason to think that they were *both* sensory.

To fully appreciate this last point, we may identify two ways to model the relationship between agentive experiences and experiences of passivity (cf. de Vignemont's (2011) discussion of the relationship between the sense of ownership and the sense of disownership). On the first model, experiences of passivity derive from psychological processes that are independent of those responsible for agentive experiences. If this were the case, then facts about experiences of passivity would not straightforwardly entail facts about the nature of agentive experiences. On this model, it may be that the psychological process responsible for the former is sensory in nature, and that responsible for the latter is not. And until we have independently determined that this

is not, in fact, the correct model, we are not licensed to make any inferences from claims about experiences of passivity to claims about agentive experiences.

On a second way of modeling the relationship between experiences of passivity and agentive experiences, they are grounded in the same psychological process, such that depending on its outcome, it may yield either type of experience. This possibility makes it more likely that, if experiences of passivity are sensory, then so too are agentive experiences, since one and the same psychological process is more likely than not to produce outputs of one and the same mental type. But again, until we have independently determined that these two types of experience *are* generated by one and the same psychological process, we are not licensed to draw any inferences from claims about one to claims about the other.

Perhaps the most pressing difficulty with Bayne's argument, however, is his claim that positing a lack of agentive experiences yields explanatory power for understanding the reports of individuals with anarchic hand syndrome. Whatever merit an explanation that posits a lack of agentive experiences may have, it supports the existence of such experiences only if it is stronger than alternative explanations. This is not the case, however. Indeed, a more straightforward, and equally strong, if not stronger, explanation is available for what underlies denials of action authorship in anarchic hand syndrome, which makes no appeal to a lack of agentive experience, nor a presence of experiences of passivity, for that matter. Individuals with anarchic hand syndrome typically recognize that they are unable to inhibit the movements of their anarchic limb, as well as aware that what their limb does conflicts with what they consciously intend to do, and that the movements of their limb are often strongly reactive to environmental triggers. But if so,

then there is no puzzle as to why they deny that they are the agents behind their anarchic movements—they do so because they are aware of lacking intentional control over these movements. Neither experiences of passivity, nor the absence of agentive experiences play a role in this equally credible alternative explanation of the psychology underlying the reports of anarchic hand patients. And if so, then we lack any reason to view such experiences as indispensible explanatory posits.

§3.2 The Argument from Cognitive Impenetrability

Another type of argument sometimes proffered for the existence of SAEs appeals to features of agentive awareness that are allegedly characteristic of perception. For instance, Bayne (2011b) argues that agentive awareness is sensory awareness on the grounds that such awareness is cognitively impenetrable, and cognitive impenetrability is a hallmark of perception. He writes, " [p]erhaps judgment can penetrate agentive experience under some conditions, but it seems clear that on the whole agentive experience exhibits the kind of doxastic impenetrability that is characteristic of perception" (p. 499).

As evidence for the cognitive impenetrability of agentive experience, Bayne (2011b) focuses once again on the case of anarchic hand syndrome. He reasons that if an anarchic hand patient were to come to believe that the actions of her anarchic limb were, in fact, her own, this would not suffice to recover her agentive experience. Bayne urges:

> The patient can no more restore the missing experience of agency by forming the belief that her anarchic actions are truly her own than you or I can correct our

visual experiences of the Müller-Lyer illusion by forming the belief that the lines in question are equal in length. (p. 498)

While this may be a tempting argument, there are good reasons to dismiss it. First, Bayne (2011b) presents no support for the claim that an anarchic hand patient would continue to lack agentive awareness after coming to believe that she is the agent of her anarchic movements. It may be instead that were she to form the relevant belief in a suitable way, she would thereby be agentively aware.[13] In denying this possibility without offering any independent reason to do so, Bayne is begging the very question at issue.

Moreover, there is reason to doubt that if some mental state is cognitively impenetrable, it must therefore be a sensory state.[14] Consider linguistic intuitions of grammaticality, which are largely considered to be types of *judgment* (e.g., Devitt, 2010). Linguistic intuitions are considered empirical data for or against classifying sentences of language as grammatical. But in some cases, they may be overridden by theoretical considerations, or an individual speaker's intuitions may not converge with those of the linguistic community to which she belongs. So a sentence that, according to one's linguistic intuitions, seems ungrammatical, may, in fact, be grammatical by one of these other standards. Now suppose that the individual with the conflicting intuition is moved by these other considerations to believe that the sentence in question is grammatical,

[13] Indeed, I will present an account of agentive awareness along these lines in Chapter 5.
[14] There is certainly considerable dispute that the converse holds—i.e., that if some form of awareness is sensory, then it must be cognitively impenetrable. For example, MacPherson (2012) argues that there is at least one case of color perception in which one's beliefs about the typical colors of objects affects the colors that those objects appear to have, and that the best interpretation of this result that appeals to cognitive penetration of color perception.

despite her persisting intuition. Here the intuition is cognitively impenetrable in the same way that Bayne supposes the anarchic hand patient's agentive experience to be cognitively impenetrable—a directly contradictory belief is not enough to overturn it. But this is certainly no reason to conclude that the linguistic intuition in question is a sensory state. In the anarchic hand case, it might be that the individual comes to believe that she is the agent of what the anarchic limb does, but that this belief gets "held in check" by a stronger belief, based on other considerations, that she is not the agent, such that it does not sway her overall assessment of the situation (cf. Armstrong, 1968; Byrne, 2009). Having the deliverances of some form of awareness persist despite the presence of inconsistent beliefs does not imply that the awareness in question is sensory in nature, as Bayne maintains.

In this section I have disarmed two leading arguments in favor of positing SAEs. In the absence of further arguments in support of their existence, and the lack of any clear introspective access to the sensory nature of such experiences, the case for positing them becomes increasingly thin. Of course, none of this yet establishes that there are no SAEs, just that we do not, on the basis of these arguments, have good reason to think so. I move on now to advancing positive considerations against the existence of SAEs.

§4 The Case Against SAEs

Sensory experiences do not exist in a vacuum; they are the outputs of some sensory modality or other. When it comes to SAEs, then, there must be some associated modality as well. As mentioned, there are two different strategies for identifying the relevant modality: reductive and non-reductive. On a reductive strategy, agentive experiences are assimilated to sensory experiences supported by one or more of the

familiar sensory modalities (vision, audition, olfaction, gustation, tactition, and proprioception). Non-reductive strategies, on the other hand, treat agentive experiences as *sui generis* experiences that exist within a *novel* sensory modality dedicated to sensing our own agency.

In this section, I will consider the leading versions of each type of strategy. While taking different approaches, the accounts I evaluate here share in common the core idea that SAEs are intimately connected to the very mechanisms underlying fine-grained, sensorimotor control. Indeed, they both adopt the comparator model (CM) as the theoretical framework within which to understand sensorimotor control—as will be familiar from Chapter 1. They then seek to account for SAEs in terms of specific components of this model.

For present purposes, the important component of the CM is the so-called forward model. Recall that prior to the performance of an action, the forward model is thought to take as input (i) a copy of the motor command generated by the motor system, which specifies fine-grained properties of the bodily movement to be performed, and (ii) the estimated current state of the central nervous system. On the basis of these two representations, it outputs predictions of the sensory consequences of the motor command both with respect to the properties of the action itself (e.g., those associated with moving one's finger to flip a switch), and the properties of the *effects* of the action (e.g., those associated with the lights going off). These predictions are then compared with the sensory feedback from the action and its consequences to yield a match or a mismatch (Blakemore & Frith, 2003; Pacherie, 2008). This information subsequently aids in the overall control of the action. As we shall see, on both accounts I will now evaluate, it is

this comparison process—between the forward model prediction and sensory feedback from the movement and its consequences—that is the purported locus of SAEs.

§4.1 Against a Reductive Account of SAEs

If SAEs are reducible to experiences in one or more of the familiar sensory modalities, an attractive candidate modality is proprioception—the sense of our limb positions and movements by way of receptors in our joints, muscles, ligaments, and skin.[15][16] After all, it is through vision and proprioception that we primarily sense our actions. But though we do not always see what our bodies are doing, we almost always feel what they are doing through proprioception, so that leaves proprioception as the dominant action-sensitive modality.[17] Perhaps, then, SAEs are to be identified with proprioceptive experiences.

At first blush, a proposal on which SAEs are proprioceptive experiences seems promising. After all, proprioceptive experiences are often described as being thin and recessive in the same way that agentive experiences are often thought to be. As O'Shaughnessy (1998) notes, "prioprioception is attentively recessive in a high degree, it

[15] I set aside mental actions such as mentally adding two plus three or imagining a pink elephant. It is worth noting, however, that sensory approaches to agentive awareness have difficulty accounting for such actions unless they adopt a sensory model of introspection, which faces many difficulties on independent grounds (see Rosenthal, 2004).

[16] This is not meant to be an exhaustive characterization of the properties that proprioception allows us to discriminate among, but one that is relevant for present purposes. A more comprehensive list would include: pressure, temperature, friction, balance, posture, and muscular fatigue (see Bermúdez, 1998, p. 132 - 133).

[17] This should not be understood as the claim that proprioception is richer or more robust than vision. Indeed, there is ample evidence that vision typically dominates over proprioception (e.g., Botvinick & Cohen, 1998; de Vignemont, forthcoming; Marcel, 2003).

takes a back seat in consciousness almost all of the time" (p. 175).[18] The reason that this recessive profile is common to both SAEs and proprioception may be that SAEs just are types of proprioceptive experience.

In addition, as I highlighted in Chapter 1, agentive awareness, whether sensory or not, constitutes a form of self-awareness—it is awareness of oneself as performing an action—and proprioceptive experiences, many have urged, are intimately related to self-awareness. In particular, some have held that proprioceptive experiences either themselves display, or provide the basis for thoughts that exhibit, immunity to error through misidentification relative to the first-person pronoun (IEM).

Such immunity has typically been discussed in the context of first-person thoughts or utterances of the form 'I am F' (where 'F' refers to some property) (cf. Shoemaker, 1968; Wittgenstein, 1953/2001). When IEM holds, it is maintained that it makes no sense to assert that someone is F on the basis of certain information, and yet to question whether it is oneself that is F. This is in contrast to other superficially similar cases. For example, suppose that I buy a lottery ticket and a friend tells me that someone has won the lottery. In this case I might correctly think or assert that someone has won the lottery, and it makes perfect sense to wonder whether it is I who have won the lottery. If I were to think or say that I have won the lottery, in this scenario, my thought would not be IEM. Another case, this time borrowed from Shoemaker's (1968) classic treatment of IEM: Suppose I am tangled up in a wrestling match and I see that someone is bleeding.

[18] But see Montero (2010), who notes that proprioceptive experiences, though they may be *typically* recessive because they are typically not attended to, may not be any more so than experiences in other modalities when they are not attended to.

I might correctly think or say that someone is bleeding, while justifiably wondering whether it is I who is bleeding.

Now contrast these examples with a case such as the following, which *does* involve IEM: I come to be aware of myself as being in pain in the normal way and I have the thought that I am in pain. Many have found it plausible to think that in this scenario, there is no question as to whether it is I who am in pain. In light of these types of case, Wittgenstein (1953/2001), for example, distinguishes between two uses of the first-person pronoun:

> There are two different cases in the use of the word 'I' (or 'my') which I might call 'the use as object' and the 'use as subject' […]. The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error […]. On the other hand, there is no question of recognizing a person when I say I have a toothache. To ask "are you sure that it's you who have pains?" would be nonsensical. (p. 66 – 67)

Tying all this back to the present discussion, following Evans (1982), many have stressed that IEM applies in an especially salient way to proprioceptive awareness, holding that if one is aware via proprioception that one's arm is bent, for example, one cannot then be mistaken as to whether it is one's *own* arm that is bent. In this way, a link might be thought to be neatly forged between proprioception and self-awareness, that would seem to make proprioceptive experiences ideal candidates for agentive experiences, which always pertain to *oneself* as an agent (see also Bermúdez, 1998; de Vignemont, 2012).

While this proposal is attractive, there is a significant problem with it. In some cases, agentive awareness diverges with respect to bodily movements that are the same; one may be aware of the same type of bodily movement as one's own action in one instance, and not in another.[19] Consider, for example, the findings of the neurosurgeon Wilder Penfield, who in the 1950s electrically stimulated the motor cortex of his patients during surgery. This led these individuals to make bodily movements that were "smooth… involving coordinated sequences of the operation of multiple muscles, which looked to have the character of voluntary actions, at least from the outside" (Wegner, 2002). And yet the patients denied authorship of their movements, making remarks like: "I didn't do that. You did" (Penfield, 1975).[20]

The case of Penfield's patients is not the only one a proponent of the present view must worry about. Anarchic hand syndrome provides another source of examples in which agentive awareness diverges with respect to type-identical bodily movements. In these scenarios, individuals perform what appear from the outside to be routine actions

---

[19] Such cases have also been appealed to in the literature on action theory. For example, Ginet (1990) writes: "The experience of *voluntary* exertion is significantly more than the mere kinesthetic perception of the exertion. I could kinesthetically feel my arm exerting force, in just the same way it does when I voluntarily thrust it forward, without experiencing this exertion as something I control, as my voluntary doing. I could experience it as something that just happens to me, unconnected with my will, while at the same time experiencing the exertion of the bodily part as just like one I might have produced voluntarily. The voluntariness of the experience of voluntary exertion is *a further* part of it, distinct from the perceptual part, an aspect that would be more conspicuous by its absence than it is by its presence" (p. 27).

[20] Marcel (2003) discusses results from an other electrostimulation happening around the same time by Hécaen et al. (1949). In this case, the patients' central thalamic nucleus was stimulated, which caused the patient to clench and unclench their fists. Marcel (2003) reports that the patients "'asserted the willful character of such induced movements' and described them as actions, saying that the movements had a goal-directed and voluntary character, but had no idea why they had made them" (p. 73). This is, of course, a different outcome than in the Penfield cases, but since a different area of the brain was stimulated, it does not undermine Penfield's own results.

insofar as the bodily movements involved are concerned, but which are nonetheless emphatically disowned by the individual.

For an account on which agentive experiences are identified with proprioceptive experiences, these cases are difficult to accommodate. Penfield's patients and individuals with anarchic hand syndrome perform bodily movements that are indistinguishable from the third-person from bodily movements constituting actions. And if so, then the proprioceptive experiences accompanying them are arguably also no different than those accompanying typical actions. But then there is no way to explain, on the present account, why agentive awareness is absent for Penfield's patients and those confronted with the movements of their anarchic hand. These kinds of cases thereby cast significant doubt on the claim that agentive experiences are properly construed as proprioceptive experiences.

One might object to this line of reasoning on the grounds that bodily movements of the same type need not yield corresponding proprioceptive experiences of the same type. Consider, by analogy, well-known color context effects in visual perception. Here, one and the same stimulus might appear to be different shades of color depending on adjacent stimuli or the background against which it is presented (see Albers, 2006; Hardin, 1993, 24; Persuh and Ro, 2011 for a striking demonstration of this in non-conscious vision). So the same color stimulus may result in different color experiences. Perhaps proprioceptive experiences corresponding to bodily movements are similarly modulated depending on the context in which a particular bodily movement unfolds. If so, then given two bodily movements A and B of the same type, the proprioceptive experiences corresponding to A may be agentive, while those corresponding to B are not.

Indeed, a proposal by Frith might help fill in the picture further here. In a series of clever experiments, Frith and his colleagues have shown that the sensory effects of our actions are attenuated relative to the effects of our non-actions (Blakemore, Frith, & Wolpert, 1999; Blakemore, Goodbody, & Wolpert, 1998; Blakemore, Wolpert, & Frith, 2000).[21] Attenuation may be understood as a comparatively lower degree of subjective intensity along one or more of a sensory quality's dimensions, e.g., the loudness of an auditory sensory quality. In one such experiment, participants were asked to apply a touch to the palm of their hand using a metal rod with a foam attachment ("self-generated" condition). In a second condition ("externally generated"), the experimenter applied the touch to the same spot using the rod. The participants were asked to rate how tickly, intense, and pleasant the touch felt in each condition. Participants consistently rated the self-generated touch to be less tickly, intense, and pleasant than the externally-generated touch. In other words, participants judged the effects of their own actions to be attenuated relative to effects that were not the result of their own actions (Blakemore, Frith, & Wolpert, 1999).

Results from neuroimaging studies also suggest a difference between the self-generated touch and externally-generated touch in the self-tickling paradigm, showing more activation in somatosensory cortex for the externally-generated touch than the self-generated touch (Blakemore, Wolpert, & Frith, 1998, 1999; Hesse, Nishitani, Fink, Jousmaki, & Hari, 2010).

[21] This phenomenon is thought to help explain why it is that we cannot tickle ourselves, but other people can tickle us.

The explanation widely offered for these findings is that, since the self-generated touch is the result of the participant's own action, a forward model prediction based on a copy of the relevant motor command is formed and then compared against the sensory consequences of the action, i.e., the tactile sensation (Blakemore, Wolpert, & Frith, 2000). Since there is a match between the two, the sensation is thereby attenuated or dampened. In the case of the externally-generated touch, however, the participant is being passively touched, and forms no motor command, and so no corresponding forward model prediction. There is therefore no comparison, and no attenuation that takes place.

Schizophrenic individuals with delusions of control were also tested on this task in a separate experiment. Recall that in delusions of control, schizophrenic individuals experience many of their actions as alien to them—they lack agentive awareness. For example, in a characteristic report associated with such delusions, the schizophrenic individual PH complains that, "[m]y fingers pick up the pen, but I don't control them. What they do is nothing to do with me" (Frith, 2007, p. 108). Now, importantly, it has been found that schizophrenic individuals with delusions of control appear *not* to experience sensory attenuation, at least to the same degree as healthy individuals. Blakemore, Smith, Steel, Johnstone, & Frith (2000) asked diagnosed schizophrenics with and without delusions of control, as well as healthy controls, to perform the same self-tickling task. They found that those with delusions of control did not rate externally-produced tactile stimuli as more "tickly" than those resulting from self-generated movements. By contrast, the schizophrenic individuals *without* delusions of control, and the healthy control participants *did* rate them as more "tickly". Similar studies have also found evidence of a lack of sensory attenuation for action effects among schizophrenic

individuals (Lindner, Thier, Kircher, Haarmeier, & Leube, 2005; Shergill, Samson, Bays, Frith, & Wolpert, 2005).

On the basis of these findings, some theorists have suggested that sensory attenuation, tied to the workings of the forward model, can be invoked to explain delusions of control in schizophrenia.[22] The idea is that *because* the sensory experiences accompanying their actions are not attenuated for schizophrenic individuals, their actions feel instead like passive movements. Regarding the case of PH, for example, Frith (2007) offers a diagnosis along just these lines:

> I believe we can now achieve some understanding of PH's experiences because of what we have discovered about the brain. In our normal state we are hardly aware of the sensations that occur whenever we move. This is because our brain can predict these sensations and suppress our awareness of them. But what would it be like if something went wrong with the prediction and we became aware of the sensations? Normally I am only aware of the sensations when someone else moves my hand. Such a brain abnormality could explain why PH feels as if her arm is being moved by someone else. She is abnormally aware of her bodily sensations when she moves her hand. For her it really does feel as if someone else were moving her hand. (p. 109)

At first blush, this proposal seems to help get around the problem raised by the Penfield and anarchic hand cases, since it appears to lend support to the idea that the same bodily movements need not be accompanied by the same proprioceptive experiences, and to give an explanation as to why. Depending on whether a forward

[22] I will take issue with this claim in Chapter 3, but I will not challenge it here.

model prediction is formed and a match registered at the relevant comparator, in one case proprioceptive experiences might be attenuated, and in the other case not, yielding different sensory experiences despite type-identical bodily movements. On this view, agentive experiences are those proprioceptive experiences that have been successfully predicted by the forward model, and are thereby attenuated.

Upon closer examination, however, the issue is not so easily resolved. The problem is that, at most, the studies I have been discussing using the self-tickling paradigm establish that sensory attenuation applies to the sensory experiences corresponding to the *effects* of our actions, i.e., the tactile stimulus applied to the palm of the participant's hand. They do not establish that such attenuation occurs for proprioceptive experiences accompanying actions themselves, i.e., the moving of the rod. It is sometimes simply assumed that sensory attenuation occurs for actions themselves and not just their consequences. For instance, Frith (2005) writes: "I am assuming that proprioceptive feedback is attenuated during voluntary movements through forward modeling" (p. 757). To date, no behavioral experiment of which I am aware has been carried out in order to test this assumption. This is a significant shortcoming in the evidence needed to support Frith's proposal.

Though there are no behavioral results, there are, however, neurophysical results that are sometimes appealed to in order to support the claim that there is subjective attenuation of sensory experiences accompanying actions and not just action effects. Blakemore, Oakley, and Frith (2001) ran an experiment in which they had participants undergo hypnosis. In the "active movement" condition, participants were asked to move their left arm up and down. In the "passive movement" condition, a pulley system moved

the participant's left arm up and down, without any contribution from the participant. In an "active-feels-passive" condition, hypnotized participants were told that their left arm was being raised by the pulley, but in actuality, they were moving their arm themselves. In a final "rest" condition, participants did nothing.

The participants underwent PET scans in all four of these conditions. Passive movements produced *greater* activation in primary and secondary somatosensory cortex, as well as the inferior parietal lobe bilaterally, compared with active movements. This greater activation was interpreted as reflecting a lack of sensory attenuation. Interestingly, these same regions showed greater activation when the hypnotized participants were merely *told* that their active movements were passive than when they performed active movements without being told this. This may be taken to suggest that merely attributing passivity to a bodily movement is enough to induce sensory attenuation—whether the bodily movement is active or passive.

There are reasons to be skeptical, however, that this study shows that there is sensory attenuation for actions themselves, and not merely their effects. One significant worry is that this interpretation relies on the shaky assumption that greater brain activation in particular regions corresponds to greater intensity of conscious sensation. This assumption is not warranted without a far better understanding than we currently possess of how brain activation relates to conscious experience. It is methodologically dangerous to draw any conclusions about the character of conscious experience based on brain activation alone.

Second, even if we do grant this problematic assumption, another study conducted by Weiller et al. (1996) found different results. Weiller et al. had participants undergo

PET scans in two conditions: an active one in which they flexed and extended their elbows, and a passive one in which their elbows were flexed and extended by a torque motor. The experiments found the same activation in primary sensorimotor cortex during both active and passive movements. The authors conclude: "Activation was almost identical in location and amount in the primary sensorimotor cortex during both passive and active movements" (p. 108).

Finally, the effects of hypnosis on the brain are still poorly understood. The relationship between hypnosis and consciousness is even less well understood. Again, it is not methodologically sound to generalize from such cases to typical cases of action or awareness. Until these results can be supplemented with evidence from a more standard experimental setting, they should not be assigned significant weight.

If sensory attenuation does not apply to bodily actions themselves, then we are left without an underlying SAE to explain cases of agentive awareness that accompanies actions for which there are no salient effects. So, until further evidence is in about whether sensory attenuation affects the experiences corresponding to our bodily actions, and not just their effects, the proposal that agentive experiences are attenuated proprioceptive experiences cannot provide a general account of agentive experiences.

In addition, even if there *were* evidence that attenuation is exhibited by the experiences corresponding to our bodily actions themselves, these results are consistent with the causal relationship running in the other direction: actions and their effects may feel subjectively attenuated as a *result* of antecedent agentive judgments. And if so, then they cannot be the basis for such judgments, as is commonly assumed. Indeed, a recent study by Desantis, Weiss, Schutz-Bosbach, & Waszak (2012) supports just this

hypothesis. In one experiment, participants were asked to judge the loudness of an auditory tone that they believed was either self-caused or externally-generated by way of key presses. In actuality, the tones were always triggered by the participants' key presses. But even so, participants judged the tones to be attenuated when they simply *believed* that they were self-caused rather than externally caused, suggesting that the effect is driven by one's beliefs about the causal relationship between one's action and its effects. This suggests, in turn, that even if attenuation were found for actions themselves, antecedent agentive judgments likely drive the effect rather than the other way around. As such these cannot be the SAEs for which we are searching.

These difficulties make it unlikely that sensory attenuation can save a reductive account of SAEs in terms of proprioception. But one might be moved to argue here that, even if proprioception alone cannot support SAEs, proprioception *plus* vision can do so. After all, though it is true that we do not see all of our actions, we do *typically* have visual feedback from our bodily movements. Perhaps, combined with proprioception, this might yield multimodal SAEs.

Though this proposal may be attractive, it does not hold up. The sensory properties of our movements accessible by vision are even less informative than proprioception with respect to whether the movements in question are actions or not. There is no visual quality or set of visual qualities that correspond to actions as against non-actions—the Penfield and anarchic hand cases double as support for this claim as well. But if not, and if, as I have just argued, there is no such quality or set of qualities in proprioception either, it is doubtful that vision and proprioception will somehow jointly result in SAEs.

Before moving on from the reductive strategy for establishing SAEs, it is worth considering an interesting variant of the approach, defended by Prinz (2007).[23] On Prinz's view, agentive experiences are forward model predictions, which he construes as a type of sensory imagery.[24] His reason for doing so is that in order for these predictions to be compared with reafferent sensory feedback, they must have the same representational format as this feedback, so they must be sensory states as well—and if so, then the most plausible option is that they are a type of sensory imagery, which Prinz understands as the "willful reactivation" of dedicated input systems.[25]

Prinz's proposal does identify a type of sensory experience that, if the comparator model is an accurate model of sensorimotor control, accompanies our actions but not our non-actions. And in this way it may seem inviting as a candidate for SAEs. Recently, however, Mandik (2010) has argued that Prinz's account runs into difficulties. Mandik's main concern is that it does not have the resources to distinguish sensory imagery from perception with matching content. He points out that sensory imagery is distinguished from perception in that the former is "willful", i.e., a type of mental action, and the latter

[23] Though see also Prinz (2012, p. 237 - 239) in which he notes some problems for the Prinz (2007) view, and ends up abandoning it in favor of a deflationary approach to agentive awareness on which there is no experience of authorship, but simply experiences of action and their disruption.

[24] It is worth noting that Prinz's (2007) original proposal is that SAEs arise from the *match* between the forward model prediction and sensory feedback. But he revises it to exclusively rely on forward model predictions in order to accommodate cases in which agentive experiences seem to persist despite there being no sensory feedback to match with a forward model prediction, as in the case of deafferented individuals and when one is under anaesthesia. I will discuss these types of cases myself in Chapter 4.

[25] This does not seem quite right, however, since the two could be compared simply in terms of their representational content, in the way that a belief about something might be compared with a perception of that thing in terms of the respective content of each state. I set aside this criticism here, however, to focus on other aspects of Prinz's view.

is not. And according to Mandik, the way we distinguish between the two from the first-person involves agentive awareness over sensory imagery, but not over perception. But, since on Prinz's view intentions do not play a role in agentive awareness, Mandik urges, we should not be able to subjectively distinguish between sensory imagery and perception. Since we can, this is a problematic consequence of the view.

There are several ways for a proponent of Prinz's (2007) view to respond to Mandik's (2010) worry (see Mylopoulos, 2011 for discussion). For one, the presence of an intention need not be the *only* way by which we subjectively distinguish between sensory imagery and perception. There is also the Humean marker of greater vivacity and intensity that accompanies perception but not sensory imagery. In addition, Prinz himself accepts that there is no subjective difference between the sensory images that are forward model predictions and perception, so he would not likely be moved by Mandik's objection. Prinz (2007) writes:

> It is perfectly plausible that we form conscious anticipatory images of our actions before executing them. These images may be generated just a few milliseconds before movement, so they are quickly succeeded by the perception of our bodies in motion. The matching process may consist in the fact that there is no significant change in the image that immediately precedes the movement and the one that follows, so we don't have any experience of two representations being compared, but only one experience of our bodies in motion, which begins just before acting and continues on after initiating movement. (p. 342)

Here Prinz seems to be saying that, in cases where sensory images, at least of the forward model type, match in content with perception, we are unable to tell them apart

subjectively. This amounts to biting the bullet that Mandik presents, and it is not clear that this is an issue insofar as the distinction between sensory imagery supported by forward modeling and the corresponding sensory feedback from bodily movement are concerned. In other words, it is not clear that these really *are* subjectively distinguishable.

The foregoing does highlight a different problem for Prinz's view, however. Forward modeling has the function of anticipating bodily movements resulting from motor commands. In cases where things are functioning normally, forward model predictions will match sensory feedback from the bodily movement. But in some cases, in particular those in which there is an error in action execution, forward model predictions will not match the sensory feedback from bodily movement. In these cases, we should expect there to be a sensory image of the intended bodily movement, and, subsequently, a perception or sensation of an erroneous bodily movement. This does not seem to capture the phenomenology of action error, however. When we make such errors, we are not first presented with the phenomenology of having successfully performed the action we are trying to perform—we are simply aware that something has gone wrong. The proponent of Prinz's view might reply that the forward model predictions are always nonconscious in these cases, but we would need independent reason to think this—otherwise the reply is plainly *ad hoc*—and no such reason is forthcoming. Prinz's imagistic forward model predictions are thus not tenable candidates for SAEs.

If agentive experiences are sensory experiences, it is unlikely that they are sensory experiences in any of the familiar modalities—whether imagistic or perceptual. The reductive strategy fails. I move on now to consider an alternative, non-reductive

account on which SAEs are thought to arise out of a novel, previously unrecognized sensory modality.

§4.2 Against a Non-Reductive Account of SAEs

As mentioned at the outset of this chapter, recently Bayne (2011b) has defended a sensory approach to agentive awareness, proposing that agentive experiences exist, not in any of the standard sensory modalities, but in a novel sensory modality that is dedicated to registering our own agency. He writes:

> Where in the cognitive architecture should we locate such experiences? They are not located within the central cognition, nor are they located within the systems responsible for programming and executing actions, nor are they located within the high-level reaches of any of the standard perceptual modalities. Instead, such states are the products of a dedicated perceptual system (or systems). (p. 494 – 495)

More specifically, Bayne, like others, urges that this sensory system is to be located at the site of the comparator that is also responsible for sensory attenuation—what I will now call the *forward model system*. Bayne's claim is that, in cases of a match between the forward model prediction and sensory feedback, the forward model system outputs an SAE. But, importantly, the SAE in question is something distinct from any attenuated sensory experience that might subsequently follow. In this way Bayne's proposal is importantly different from that of Frith and his colleagues.

There has been much debate surrounding whether the forward model system, as just described, adequately accounts for all the cases in which agentive awareness arises and fails to arise. For example, Synofzik et al. (2008) argue that it cannot adequately

explain a range of cases, while Carruthers (2012) is more sympathetic, arguing instead that with suitable adjustments, it can do the explanatory work required of it. Wong (2012) further points out that if one is too liberal with these adjustments, as may be required for sufficient explanatory power, the question arises as to whether "the term 'comparator' is just standing proxy for whatever mechanism is responsible for the sense of agency insofar as some feedback is involved" (p. 50). These are important issues, and I will comment on them further in Chapter 3, but I put them to one side here, as I will not be evaluating Bayne's proposal from this angle. My focus instead will be to argue that the forward model system is not suitable for supporting SAEs at all, let alone explaining the range of cases in which agentive awareness allegedly arises, and that Bayne's proposal should therefore be rejected on this basis.

Bayne's account requires that we construe the forward model system as its own sensory modality. This is a difficult claim to evaluate, partly due to the challenge of specifying what counts as a sensory modality. But we need not supply necessary and sufficient conditions for being a sensory modality in order to achieve our goal. Instead, we may appeal to hallmark features of sensory modalities, determine whether the forward model system exhibits them, and evaluate on that basis whether it is credibly viewed as a sensory modality. Once we do so, we will see that Bayne's proposal is beset with difficulties.

Perhaps the most reliable feature associated with sensory modalities is that they operate by way of sense organs over which we have some degree of control, whether directly or indirectly (Keeley, 2002; Shoemaker, 1994). We have, for example, our eyes for vision, our ears for audition, our nose for olfaction, and our tongues for tasting, and

we have some degree of control over each of these. When it comes to the forward model system, however, there is no candidate sense organ over which we have any degree of control. The comparison that takes place between forward model predictions and sensory feedback is not something that we can initiate, guide, or inhibit at will. So the case for viewing the forward model system as a sensory modality seems, in this crucial respect, to break down.

Bayne (2011b) acknowledges this issue, but urges that the characterization of a sense organ as something over which we have some degree of control is misguided, since we do not have any control over the mechanisms responsible for proprioception, nociception, and vestibular sense, and yet they still qualify as sensory modalities. In light of this, Bayne suggests that a sense organ be construed instead as "a dedicated mechanism that takes as input energy of some kind and generates representations in an appropriate format, at least some of which are experiential" (p. 512). On this characterization, the mechanisms underlying proprioception, nociception, and vestibular sense would arguably qualify as sense organs and, thereby, as sensory modalities.

But even if one accepts Bayne's characterization of a sense organ, which may indeed be appropriate in the case of proprioception, nociception, and vestibular sense, the forward model system falls short. For, as Bayne himself points out, instead of taking as input raw energy, e.g., wavelengths in the case of vision, it takes psychological representations as input, namely the forward model prediction and sensory feedback. This is unlike any sensory modality with which we are familiar. Indeed, it stretches the notion of a sensory modality beyond recognition—sensory modalities have, at their core, the

function of discriminating among sensory properties of our bodies and the environment. They do not discriminate among properties of psychological representations.

Bayne (2011b) maintains that this difference in inputs among typical sensory modalities and the forward model system is not a problem. The forward model system, he suggests, should simply be viewed as a "non-basic" sensory modality, whereas the sensory modalities that take as input raw energy should be construed as "basic" modalities. Without some independent reason to bifurcate sensory modalities into basic and non-basic kinds, however, this reply lacks any motivation. In addition, it is entirely mysterious how to understand sensory properties of psychological representations in a way that is remotely analogous to the sensory properties, e.g., colors and shapes, of physical objects on which familiar sensory modalities operate. Sensory properties can be grouped into families according to their similarity and difference relations to one another. But no such grouping can be applied to properties of psychological representations. Again, to view the properties of psychological representations as sensory properties seems to distort the notion of a sensory property beyond recognition.

Finally, a further problem is that adopting Bayne's characterization of non-basic sensory modalities leads to an excessively liberal view of sensory modalities. Consider, for example, the case of arithmetic calculations. Suppose one compares two sums with respect to which one is greater, and that the output of such a comparison is a judgment to the effect that the first sum is greater than the second sum. It would seem to follow, on Bayne's characterization of sensory modalities, that this is an instance of sensory processing, and that we have a sensory modality dedicated to arithmetic. After all, the inputs are psychological representations and, if we countenance cognitive

phenomenology, which many do, then they may output representations, i.e., conscious thoughts, that are at least sometimes "experiential" in character. But a view of sensory modalities on which it follows that we have an arithmetic sense seems to have led us far astray from any informative notion of these entities.

In sum, there is good reason to deny that the forward model system is properly construed as a sensory modality. And if so, then there is reason to deny that the workings of the forward model system constitute a form of perception, and thus that it could be home to SAEs

§5 Conclusion

Agentive experiences are commonly posited in discussions of agentive awareness. But rarely is their precise nature discussed. In this chapter, I have argued against a particular view of such experiences, namely the view that they are sensory experiences. Not only do leading arguments in favor of positing SAEs fail, but there is no sensory modality or set of sensory modalities—either familiar or novel—within which to locate them.

But the considerations I have advanced in this chapter are not the only ones available against a sensory account. Indeed, some of the considerations I put forward in later chapters against alternative views of agentive awareness double as reasons against adopting a sensory approach to agentive awareness. In the next chapter, for example, I will advance general considerations against appealing to the comparator model as an explanation of agentive awareness, which apply also to Bayne's non-reductive model of SAEs. And in Chapter 4, in the process of arguing against a model of agentive awareness

that requires a match between goals states and monitoring states, I will argue that there are cases in which agentive awareness arises in the absence of any sensory feedback, which also speaks against any sensory account of agentive awareness. So these further arguments will serve to supplement the ones I have presented here.

The upshot of all this is that, if we are to countenance agentive experiences, we must look to non-sensory models to explain them. I will suggest such a model in Chapter 5, but in the meantime I move on to presenting a general challenge to views that appeal to the workings of the comparator model to explain agentive awareness.

CHAPTER 3: THE LOW-LEVEL APPROACH TO AGENTIVE AWARENESS

§1 Introduction

In Chapter 1, I characterized action control very generally as being comprised of two main functions: guidance and monitoring. I also introduced two types of action control: intentional control and sensorimotor control. One way in which these two types of control are distinct, as I mentioned there, is in terms of the types of mental state involved. Intentional control, recall, involves guidance and monitoring that is implemented by mental states bearing intentional properties—states like intentions, thoughts, and perceptions. We conceptualize our actions and the events and objects in our surroundings in virtue of being in such states. Sensorimotor control, on the other hand, involves guidance and monitoring that is implemented by states lacking intentional properties, but having representational character in terms of their sensory and motoric properties—states like motor commands and bodily and visual sensations.

I also distinguished between three types of theoretical approach to agentive awareness with respect to what aspects of action control, and therefore what types of state and what types of process, it is viewed as arising out of: (i) high-level, (ii) low-level, and (iii) integrated approaches. Recall that on a high-level approach, an individual is agentively aware of her actions in virtue of the states and processes involved in the intentional control of her action. On a low-level approach, recall that an individual is agentively aware of her actions in virtue of the states and processes involved in the sensorimotor control of her action. Finally, integrated approaches combine the commitments of high-level and low-level approaches. Such approaches to agentive awareness have it that an individual is agentively aware of her actions in virtue of the

states involved in both intentional control and sensorimotor control (e.g., Pacherie, 2008; Synofzik, Vosgerau, & Newen, 2008; Gallagher, 2007).

In this chapter, I will argue that the states and processes involved in sensorimotor control do not contribute to agentive awareness. If I am right, then both low-level and integrated approaches, which take on board this commitment, are undermined. The result is that high-level approaches emerge as the dominant contenders in theories of agentive awareness.[26]

I will focus on the leading, and arguably only, low-level account in the literature, which holds that agentive awareness arises out of the states and processes involved in sensorimotor control, as described by the comparator model (CM), in particular. Recall that the CM posits motor commands (i.e., efferent representations), forward model predictions, and sensations (i.e., sensory afferent feedback) in order to explain how it is that we are able as agents to exercise fine-grained control over our actions. More specifically, as I detailed in Chapter 1, the comparator model consists of two internal models: the inverse model and the forward model. The former outputs the motor command needed to drive the body from its current state to the intended state, and the latter outputs a prediction of the state the body will be in if the motor command produced by the inverse model is executed.

In addition, three separate comparisons are carried out between the intended state, forward model prediction, and sensory feedback from the bodily movement. With respect to agentive awareness it is the output of the first comparison, between forward

[26] In the next chapter, I will consider and reject an influential version of the high-level approach, and in Chapter 5, I will offer my own high-level account of agentive awareness.

model predictions and sensory feedback, that is typically the focus of low-level theorists (Bayne, 2011b).[27] So these are the states on which I fill focus as well. The idea, once again, is simple: When there is a match between the forward model prediction and sensory feedback, one is agentively aware of performing an action. When there is a mismatch, one is not.[28]

The chapter is organized as follows: In section 2, I lay out the evidence that has been offered in support of the low-level, CM-based approach—henceforth the CM approach—to agentive awareness. Section 3 is devoted to arguing that this evidence does not actually support the CM approach, in that the states and processes of the CM cannot consistently explain the relevant data. In section 4, I advance my positive argument against the CM approach to agentive awareness, which is that sensorimotor control of the

[27] Some theorists are more specific, urging that the comparison between the intended state and the forward model prediction underlies one's sense of *initiating* an action and the comparison between the forward model prediction and sensory feedback underlies one's sense of *control* over an action (e.g., Synofzik, Vosgerau, & Newen, 2008). But the main evidence appealed to in support of positing the former comparison is not compelling. It is that a number of studies have found that people tend to report anticipatory awareness of acting—that is, they report having started acting before any sensory feedback from the action is available (Lau, Rogers, Haggard, & Passingham, 2004; Libet, Gleason, Wright, & Pearl, 1983). Such reports, however, could just as well be based on intentions to start acting, rather than the output of any comparison involving a forward model prediction, as I will suggest in Chapter 5. I will thereby restrict my discussion to the comparison between the forward model prediction and sensory feedback, for which there is arguably more substantive evidence.

[28] This way of putting the relationship between low-level states and agentive awareness leaves it open as to whether the contribution on the part of low-level states is supposed to be direct or indirect. If the contribution is direct, then at least part of the content of these states enters into agentive awareness. If the contribution is indirect, then low-level states merely influence the content of agentive awareness at a higher-level. (This distinction was helpfully suggested to me by Elisabeth Pacherie). My critique in what follows of the low-level approach should apply regardless of whether the contribution of low-level states to agentive awareness is understood as direct or indirect, so I remain neutral on this point.

type that the CM is meant to capture dissociates from agentive awareness. I end by considering some objections and replies to this challenge.

§2 The Case for the CM Approach to Agentive Awareness

I wish to stress that I am not here critiquing the CM as a model of sensorimotor control. Indeed, there is evidence that predictive forward modeling, which is a central component of the CM, plays an important role in sensorimotor control.[29] For example, people will adjust their grip force in anticipation of the weight of the object they are about to lift (Flanagan & Wing, 1997; Wing, Flanagan, & Richardson, 1997).

The case for the CM approach to *agentive awareness*, however, is on shakier ground. This case, too, has been built largely on a foundation of empirical data. There are at least three main strands of evidence that are adduced in favor of the CM approach. Together, they make up an impressive body of empirical work that offers much insight into the nature of agency and its connections with conscious experience. As I will argue, however, it does not speak in favor of the CM as an explanation of agentive awareness. In this section, I describe the relevant empirical results.

§2.1 Sensory Attenuation Studies

Perhaps some of the earliest and most frequently cited results in favor of a CM approach to agentive awareness has emerged from the sensory attenuation studies that were introduced in Chapter 2 (Blakemore, Wolpert, & Frith, xxxx; see also Weiskrantz, Elliott, & Darlington, 1971). Recall that in the standard paradigm, participants use a device—attached with a foam-tipped rod—to touch their right palm ("self-generated"

[29] Recently, some have even tried to extend the CM approach to account for language production and comprehension (Pickering & Garrod, 2013). See Mylopoulos and Pereplyotchik (2013) for a critical commentary surrounding the empirical motivation for this project.

condition). In another condition, the experimenter instead controls the device to touch the participants in the same place on their right palm ("externally-generated" condition). After each trial, participants rate on a scale the "tickliness", "intensity" and "pleasantness" of the tactile stimulus.

The key result, first reported in Blakemore, Goodbody, & Wolpert (1998) and subsequently replicated (Blakemore, Frith, & Wolpert 1999; Blakemore, Smith, Steel, Johnstone, & Frith, 2000), is that, on average, participants judge the self-generated tactile stimuli to be *less* tickly, intense, and pleasant than those that are externally generated. Blakemore and colleagues refer to this general phenomenon as *sensory attenuation*.

Importantly, Blakemore, Frith, & Wolpert (1999) found that sensory attenuation is sensitive to fine-grained details of the spatial and temporal properties of the tactile stimulus. They included trials in which they manipulated either the spatial trajectory or the timing of the tactile stimulus. The participants' tickliness ratings increased for these trials i.e., the sensory attenuation effect decreased, as the delay between the movement of the left hand and the tactile stimulus on the right hand increased from 0 to 300 ms. The greater the delay, the more "tickly" the tactile stimuli were judged to be. Indeed, self-generated tactile stimuli that were applied after delays of 200 and 300 ms were judged to be just as "tickly" as externally-produced tactile stimuli. The sensory attenuation effect was canceled out in these cases. The participants' tickliness ratings also increased as the spatial trajectory between the movement of the left hand and the tactile stimulus on the right hand increased from 0 to 90 degrees.

These results have been widely taken to suggest that sensory attenuation is an effect of forward modeling. The idea, as I explained in Chapter 2, is that when the

sensory consequences of our bodily movements are successfully predicted by the forward model, they are subjectively dampened. As Blakemore, Wolpert, and Frith (1999) put it: "… an estimate of the tactile feedback from the hand movement is made and, when congruent with the actual sensory consequences, this estimate is used predictively to attenuate the percept of the tactile stimulus" (p. 457). Since forward model predictions accompany only our own actions, the tactile stimuli in the self-generated condition are predicted and thereby attenuated, while those in the externally-generated condition involving the actions of another agent are not. And as the temporal or spatial properties of the predicted consequence, i.e., the tactile stimulus, are unexpectedly altered in the perturbation conditions, the forward model prediction, which is very precise, becomes inaccurate, so sensory attenuation decreases even for the effects of our own actions.

The case for relating agentive awareness to forward modeling makes a central appeal to the behavior of schizophrenic individuals with delusions of control on self-tickling tasks. Recall from Chapter 2 that schizophrenic individuals with delusions of control appear not to experience sensory attenuation, at least to the same degree as non-schizophrenic individuals (Blakemore, Smith, Steel, Johnstone, & Frith, 2000). Many have thereby suggested that sensory attenuation, based on the workings of the forward model, can be invoked to explain delusions of control in schizophrenia. More specifically, an impairment in forward modeling leads to a lack of sensory attenuation in these individuals, and as a result no agentive awareness. As Blakemore (2003) writes:

> An impairment in such a predictive system [i.e., the forward model] could cause a lack of attenuation of the sensory consequences of self-produced actions, which would therefore be indistinguishable from externally generated sensations. This

would result in the interpretation of one's own movements as being externally caused—a delusion of alien control. (p. 649)

On this view, sensory attenuation is the result of forward modeling, and agentive awareness results from sensory attenuation. So there is a connection forged between forward modeling and agentive awareness.

§2.2 Action-Monitoring Studies

Another popular experimental paradigm sometimes used to make the case that agentive awareness depends in some way on forward modeling is the action-monitoring paradigm (e.g., Daprati et al., 1997; Farrer, Bouchereau, Jeannerod, & Franck, 2008; Farrer et al., 2004; Farrer et al., 2003; Fourneret & Jeannerod, 1998). In this paradigm, various methods are used to distort visual feedback such that it no longer corresponds to the action that the individual is performing.

In a typical study using this paradigm, a participant is asked to perform an action, e.g., move a joystick (Farrer et al., 2003) or trace a straight line (Fourneret, Paillard, Lamarre, Cole, & Jeannerod, 2002) while they are seated facing a virtual image of a virtual hand performing the same action. The virtual image is, on some trials, distorted with respect to its spatial or temporal properties so that it no longer corresponds to the action being performed by the participant. After each trial, participants are asked whether the movement they see is their movement or not. The more drastic the distortion, the more likely participants are to report that the movement they see is not their own.

It is commonly thought that the reports the participants in these studies make with respect to whether the movement they see corresponds to the one they are performing is

based on the outcome of a comparison between the forward model prediction and the visual feedback they receive. And once again, schizophrenic individuals with delusions of control perform differently than healthy controls on this task. In an early study, Daprati et al. (1997) had healthy participants, schizophrenics with delusions of control, and schizophrenics without delusions of control perform the action-monitoring task. They found that schizophrenics with delusions of control made more errors on the task than each of the other groups. They were more likely to report that the distorted sensory feedback corresponded to their own movements than schizophrenic individuals without delusions of control and healthy controls.

A more recent study by Synofzik, Thier, Leube, Schlotterbeck, & Lindner (2010) even suggests that the performance of schizophrenic individuals with delusions of control on these tasks is correlated with the severity of their symptoms. In one experiment, schizophrenics with delusions of control and healthy control subjects were asked to make straight "out-and-back" pointing movements on a board with their right index fingers as quickly as they could. Participants could not see their movements. Rather, they were seated facing a monitor that displayed visual feedback that was rotated to varying degrees in a direction either clockwise or counterclockwise relative to the actual movement. After performing each movement, participants were asked to report the direction (clockwise or counterclockwise) that they judged the visual feedback to be rotated. The experimenters then calculated the just noticeable differences for feedback deviation.

Schizophrenic individuals were less accurate at detecting feedback manipulations than healthy controls. They demonstrated larger just noticeable differences than controls, and with significantly higher variability. Interestingly, the size of individual patients' just

noticeable differences significantly correlated with their scores on a test measuring the severity of their delusions of control. The higher the score on this test, the less accurate they were at detecting distortions in visual feedback.

Here, too, the performance of schizophrenic individuals in these studies is thought to reveal an impairment in forward modeling (e.g., Synofzik, Thier, Leube, Schlotterbeck, & Lindner; Farrer et al., 2004). And since the worse their delusions of control, the worse their performance on this task, the alleged breakdown in forward modeling is thought to be the cause of the accompanying breakdown in agentive awareness.

§2.3 Intentional Binding Effect

Yet another empirical result that is appealed to in support of the CM approach to agentive awareness is the *intentional binding effect* (Haggard, Clark, & Kalogeras, 2002), also briefly introduced in Chapter 2. The experimental paradigm typically used to establish this effect is set up as follows: First, in a single-event baseline condition, participants are asked to report the time at which they become aware of the onset of one of four events: (i) their own action of pressing a key, (ii) a muscle twitch induced by TMS applied to motor cortex, (iii) a 'click' of a sham TMS applied to parietal cortex, and (iv) an auditory tone presented on its own. Next, in a second condition, participants are asked to once again report the time at which they first become aware of a key press, a muscle twitch induced by TMS, and a 'click' of a sham TMS, except this time each of these events is followed 250 ms later by an auditory tone.

The crucial finding is that in the case where individuals were asked to subjectively judge the time interval between their key press and the auditory tone, their

awareness of the key press was shifted later in time relative to the baseline condition, and their awareness of the tone was shifted relatively earlier in time. In other words, there was a subjective time compression between the key press and the tone. On the other hand, their awareness of the involuntary muscle twitch was shifted earlier in time relative to the baseline condition, and the auditory tone that followed was shifted relatively later in time (see Haggard, Clark, & Kalogeras, 2002; Haggard & Cole, 2007; Tsakiris & Haggard, 2003).

Since actions, like the participants' key presses, are thought to involve forward model predictions, and involuntary movements, like their TMS-induced muscle twitches, are not, the intentional binding effect has been interpreted as evidence that forward model predictions influence subjective awareness in the form of an altered experience of timing that brings together actions and their effects. This is thought to facilitate learning of salient action-effect pairings. But it is also interpreted as reflecting an "implicit" measure of one's agentive awareness (e.g., Moore, Wegner, & Haggard, 2009), where 'implicit' refers to the awareness being expressed behaviorally rather than in explicit verbal reports. Indeed, some go so far as to claim that intentional binding is responsible for *generating* the sense of agency. For example, Aarts, Custers, & Wegner (2005) write: "As Haggard et al. (2002) show, under such goal-directed circumstances the perception of the action and effect bind together to create a sense of personal authorship" (p. 455).

These three strands of empirical work are appealed to time and again in the literature in order to justify the CM approach to agentive awareness. Individually, sensory attenuation, the results from action-monitoring studies, and the intentional binding effect are thought to help establish a necessary connection between forward modeling and such

awareness. Though this fascinating work is doubtless important for understanding other aspects of action and awareness, in the next section, I argue that the case in favor of linking it with agentive awareness does not hold up. These results cannot be appealed to in support of a CM approach to agentive awareness, as they so often have been.

§3 Evaluating the Case for the CM Approach to Agentive Awareness

There are two general lines of reasoning, via the empirical results just discussed, that are employed to make the case that agentive awareness depends in some way on forward modeling, and is thus a low-level affair. The first focuses on delusions of control in schizophrenia, and starts with the hypothesis that these are caused by a breakdown in forward modeling. Next is the claim that sensory attenuation and performance on action-monitoring tasks also involve forward modeling. So, the reasoning goes, given that schizophrenic individuals with delusions of control are impaired on these tasks compared with healthy controls, as would be expected if their forward modeling were impaired, this evidence in favor of the hypothesis being tested: delusions of control in schizophrenia are due to breakdowns in forward modeling. And if so, then *intact* forward modeling is what gives rise to agentive awareness.

The second line of reasoning is more simplistic and runs as follows: Sensory attenuation, the intentional binding effect, and judgments in action-monitoring tasks, are either features of or expressions of agentive awareness. Sensory attenuation and the intentional binding effect, for example, are sometimes described as qualitative components of agentive awareness—features of the subjective experience of such awareness. And judgments in action-monitoring tasks are sometimes construed as

expressions of agentive awareness over the relevant movements (e.g., Farrer et al. 2003). If sensory attenuation, the intentional binding effect, and judgments in action-monitoring tasks all depend on forward modeling in some important way, as many suppose, then, the argument runs, agentive awareness itself also depends on forward modeling in some important way.

In order to dismantle these argumentative tactics, it will be enough to challenge the key premise that is common to both, i.e., that sensory attenuation, performance on action-monitoring tasks, and the intentional binding effect all involve forward modeling. Below I argue that an explanation in terms of forward modeling is not mandated by the data, and, in some cases, it is not even a workable explanation of the data.

§3.1 Divorcing Sensory Attenuation From Forward Modeling

When it comes to explaining sensory attenuation, the going hypothesis is that it is the result of sensory feedback, in this case, a tactile stimulus, being predicted by the forward model and thereby subjectively dampened. But an equally credible explanation is that the effect is driven instead by what are sometimes referred to as *cognitive expectations* (Knoblich & Repp, 2009) about what will occur—and these are simply learned predictions about upcoming events.

Forward model predictions are not cognitive expectations. They have a different functional role within one's mental economy, as they are formed on the basis of an efference copy and current state of the body. As Hohwy and Frith (2004) emphasize when characterizing forward model predictions:

> It is important to notice that the notion of 'prediction' used here is different from the more usual notion of conscious prediction (as in, e.g., I predict that it will

rain). The computations of the forward model happen at a subpersonal level… (p. 185)

Cognitive expectations, on the other hand, are formed on the basis of background and occurrent beliefs, desires and intentions, as well as perceptual information. These expectations, which need not be conscious, are ubiquitous; they are based on observed regularities, and help us to coordinate many of our daily actions. I expect that turning on the stove will cause the water in the kettle to boil shortly, and plan my actions in the interim and thereafter accordingly. Once associations between an action and certain effects have been learned, we are able to predict and thereby anticipate these effects. And once we form cognitive expectations with respect to an effect, arguably the fewer attentional resources we will devote to it, saving these resources for more salient stimuli and events. This should apply to the self-tickling results as well. The cognitive expectation regarding the tactile stimulus will be more accurate in the condition in which it is self-generated than in the externally-generated condition—it will be more predictable in the former case than in the latter case. And, on the present proposal, this will presumably lead to *more* attention being devoted in the externally-generated condition, less attenuation, and a relatively higher "tickliness" rating as a result.

Blakemore et al. (1999) take issue with this alternative interpretation of their results. They argue that since a robotic arm was used to create the tactile stimulus in the externally-generated condition, it was no more predictable at the level of cognitive expectations than in the internally-generated condition. They write:

… the stimulus was highly controlled and highly

predictable in time and space in the externally produced tactile sensation

condition. Therefore, the lack of stimulus predictability could be ruled out as a factor responsible for increasing tickliness rating in this condition. (p. 552)

But what I am suggesting here is not that there would be no predictability in the externally-generated condition by way of cognitive expectations, but rather that there would be relatively *less* predictability. In the self-generated condition, the individual has her own intentions, which help guide the movements of the robotic arm, to inform her cognitive expectations. This is not the case in the externally-generated condition. So, even if the onset of the stimulus were highly controlled and predictable in general, the movements of the robotic arm would be less predictable in the external condition than in the self-generated condition.

In general, sensory attenuation may be understood, not as the result of forward model prediction, but as the result of devoting more attentional resources to externally-generated stimuli as opposed to self-generated stimuli. Indeed, this falls nicely in line with a classic model of attention known, conveniently, as the *attenuation model* (Treisman, 1964), on which unattended stimuli are attenuated relative to attended stimuli, and attention is allocated based on the salience of a stimulus. In the case of the self-tickling studies, presumably in both cases the tactile stimulus is attended to by the participant. The point, however, is that it is *less* attended in the self-generated case, and this could reasonably be expected to lead to a *relative* attenuation effect like the one revealed in these studies.

Other credible explanations of the result that do not appeal to forward modeling are also available.[30] Recall from Chapter 2 the study by Desantis, Weiss, Schütz-Bosbach, & Waszak (2012), which supports the view that sensory attenuation is due to antecedent beliefs about authorship. The participants were asked to judge the loudness of an auditory tone that they believed was either self-caused or externally-generated by way of key presses. In reality, the tones were always triggered by the participants' key presses. But even so, participants judged the tones to be attenuated when they believed that they were self-caused rather than externally caused.

Indeed, not only do these results suggest an alternative explanation that does not appeal to forward modeling, but they actively *undermine* an explanation of sensory attenuation in those terms. In each trial of the experiment, the participant performed the action of pressing the key prior to the auditory tone. So a forward model prediction of the auditory tone would have been formed on each trial. But then the difference in the participants' attenuation judgments across trials cannot be explained by appeal to forward modeling. Another mechanism must be responsible, and the one that is suggested here appeals to cognitive expectations.

§3.2 Divorcing Action-Monitoring From Forward Modeling

[30] There is also an interesting study by Voss, Ingram, Haggard, & Wolpert (2006) suggesting that sensory attenuation results from mechanisms occurring *upstream* of primary motor cortex. In this study, TMS was applied to left primary motor cortex, which caused a delayed voluntary movement of participants' fingers. A tactile stimulation on the relevant finger during the delay period prior to movement was sensorily attenuated, suggesting that sensory attenuation relies on processes occurring prior to the dispatch of motor commands. This does not rule out the hypothesis that forward model predictions are responsible, however, since the CM allows that these are formed prior to the sending of the motor command to the spinal cord and muscles.

With respect to the action-monitoring studies, participants' judgments as to whether the movement they see corresponds to what they are doing is thought to be based on a comparison between the forward model prediction and visual feedback. If there is a match, they judge that the visual feedback corresponds to what they are doing, and if there is a mismatch, they judge that it does not. But an alternative explanation, which does not appeal to forward modeling, is that participants' judgments are based strictly on a comparison between their proprioceptive feedback and visual feedback. The forward model prediction need not contribute to the judgment at all.

Indeed, there is reason to think that this alternative explanation better captures the strategy being used by participants. Consider the performance of GL, a deafferented individual who does not receive proprioceptive feedback from her bodily movements, in another action-monitoring experiment.[31] Farrer, Franck, Paillard, & Jeannerod (2003) found that GL was "massively impaired" compared with controls in detecting distortions from visual feedback. On the assumption that GL's forward model predictions are intact, and given that she does perform actions, a reasonable explanation for this impairment is that GL cannot compare proprioceptive feedback with visual feedback, while healthy controls can do so.

A second reason to think that participants do the task by comparing visual and proprioceptive feedback is that the questions posed to the participants in these studies often invite this strategy. For example, Franck et al. (2001) asked, "Did the movement you saw on the screen exactly correspond to that you have made with your hand?" And Farrer, Franck, Paillard, & Jeannerod (2003) asked, "Is the movement you see concordant

[31] I will discuss deafferentation further in Chapter 4.

with the movement you are making?" These questions strongly suggest that participants are to compare the hand movement they feel they are performing (via proprioception) with the visual feedback presented to them. Indeed, the task could just as easily be performed using passive movements, highlighting the otiose nature of the forward model prediction for the purposes of successfully completing this task.

There are some hurdles to adopting the alternative interpretation I am presently suggesting, however. First, in the Farrer, Franck, Paillard, & Jeannerod (2003) study just mentioned, the participants in the first experiment underwent both an "active" and a "passive" condition. In the "active" condition, they were instructed to move a joystick in specific directions after which they were asked to report whether the movements they saw in a virtual image matched the movement they had made or not. But in the "passive" condition, the movements of the joystick were instead controlled by the experimenter, while the participants merely held onto it as in the active condition. The critical result is that participants performed better on the task in the *active* condition than in the *passive* condition. That is, they were more accurate at judging when the movements they saw matched the movements they had themselves made.

Since proprioceptive feedback is available in both cases, this would seem to go against the interpretation of action-monitoring results that I am proposing. And since forward modeling is available only in the active condition, for which the participants' performance was better, it would seem to weigh in favor of an explanation in terms of forward modeling. Indeed, Farrer et al. (2003) interpret this result as supporting an explanation in terms of forward modeling, proposing that "conscious monitoring of the outcome of the comparison process between the prediction of the sensory consequences

of an action and the actual consequences of such an action could also represent a possible basis for knowledge about the movement" (p. 618).

While it is tempting to go along with the authors in their interpretation, one should resist doing so. The issue is that the proprioceptive information available in the passive and active conditions also differs. In the passive condition, although there is proprioceptive information concerning limb position and movement, there is none concerning feedback from muscle activity, whereas in the active condition there is. This additional information could help participants better discriminate their own movements in the active condition by way of comparisons between proprioception and vision alone.[32]

Another threat to the alternative explanation I am proposing is that there appears to be some evidence from another study by Tsakiris, Haggard, Franck, Mainy, & Sirigu (2005) that people are better at determining whether visual feedback they receive corresponds to their own movements when those movements are self-generated rather than externally-generated, and when care has been taken to ensure that proprioceptive and visual cues are matched between the movements. But if so, then it would seem that forward model predictions may very well be playing a role in enabling the enhanced discriminations.[33]

When looking at the details of this study, however, a confounding factor reveals itself. In one condition ("self-generated"), participants were asked to press the left side of

---

[32] Even if one rejects this possible confound, and with it the conclusion participants in these studies compare proprioceptive and visual information in order to complete the tasks, an explanation in terms of forward modeling is still not mandated. For the participants in the active condition could have an advantage, not on the basis of forward modeling, but on the basis of forming intentions to move the joystick in certain ways. Indeed, this interpretation fits comfortably with the positive account of agentive awareness I will advance in Chapter 5.

[33] I am grateful to Élisabeth Pacherie for pressing this worry.

a lever with their left hand. This would lift up their right index finger, which they placed resting on the right side of the lever. In another condition ("externally-generated"), the experimenter would instead press the left side of the lever, lifting up the participant's right index finger. Within both conditions, participants either viewed their own right hand being moved or that of the experimenter—both were covered with gloves. They then had to answer, after each trial, whether the hand they saw moving was their own. The key result is that participants were better at recognizing their own right hand when its movement was self-generated than when it was externally generated.

Tsakiris, Haggard, Franck, Mainy, & Sirigu (2005) interpret this result as suggesting that so-called "efferent signals", i.e., low-level states like motor commands or forward model predictions, contribute to self-recognition in these types of tasks, and not just proprioceptive and visual feedback. But this is too quick, for there is extra proprioceptive information in the self-generated condition that is not present in the externally-generated condition, coming from the pressing of the lever with the left hand. Since, given the lever mechanism, the movement of the left hand is strictly related in timing to the movement of the right hand, this proprioceptive information adds a significant advantage for recognizing the movement of one's own right hand in the self-generated but not the externally-generated condition. Thus, we still have no convincing evidence that participants are going on anything more than proprioceptive and visual information in these action-monitoring tasks. And if so, an explanation in terms of forward modeling is not appropriate here either.

§3.3 Divorcing the Performance of Schizophrenic Individuals From Forward Modeling

So far I have not disputed an explanation in terms of forward modeling for why schizophrenic individuals perform differently than healthy controls in both the self-tickling studies and the action-monitoring studies. I have merely provided alternative explanations that do not appeal to forward modeling for the performance of healthy individuals. I turn now to arguing that forward modeling cannot account for the performance of schizophrenic individuals on these tasks.

In the sensory attenuation case, the explanandum is that schizophrenic individuals with delusions of control do not report that "self-generated" stimuli are less tickly, intense, and pleasant than "externally-generated" stimuli, as healthy controls do. Presumably, this reflects the fact that, for such individuals, self-generated stimuli are not attenuated in the typical way. And this, in turn, suggests to many that the process that is responsible for sensory attenuation in healthy individuals is disrupted in schizophrenics. The proposal we are considering is that the mechanism underlying sensory attenuation is a match between a forward model prediction and sensory feedback. If so, and if we go along with the reasoning thus far, then the thought is that in schizophrenic individuals there is no match between the forward model prediction and sensory feedback.

Why might this be? There are three possible ways that a match could fail to occur: (i) no forward model prediction is formed, (ii) an inaccurate forward model prediction is formed, or (iii) an accurate but imprecise forward model prediction is formed. I will consider each of these in turn. But first, a brief remark about the difference between the second and third possibilities is in order. Accuracy is a function of truth, whereas precision is a function of specificity. These sometimes come apart; a representation can be inaccurate while being precise, and it can be accurate while being imprecise. Suppose,

for example, that I am asked to guess what number someone is thinking of, and that the number in question is 7. I might guess that the number is between 6 and 9. If so, my guess would be accurate, but imprecise. On the other hand I might guess that the number is 8.7, and if so, my guess would be precise but inaccurate. As we will see, keeping these two properties distinct will be important for evaluating the role of the forward model in explaining the performance of schizophrenic individuals in self-tickling tasks.

Turning now to the three potential explanations just suggested for this performance, taking the first possibility, suppose that no forward model prediction is formed, so that there is no comparison at all with sensory feedback from the tactile stimulus, and so no match between the two is registered. While this would explain why sensory attenuation does not occur, there is a significant problem with this proposal. As Frith (2012) points out, given the putative role of forward model predictions in implementing sensorimotor control, an absence of such representations would likely lead to large deficits in executing actions. And, as Frith and others note (e.g., de Vignemont & Fourneret, 2004; Fourneret, Franck, Slachevsky, & Jeannerod, 2001; Leudar, Thomas, & Johnston, 1994; Trillenberg et al., 1998), schizophrenic individuals do not seem to exhibit impairments in the sensorimotor domain; their fine-grained control over their everyday actions is comparable to that displayed by healthy individuals, and there seems to be no impairment while performing the specific tasks required in the self-tickling studies.[34] It is unlikely, therefore, that there is a failure to compute forward model predictions altogether in the self-tickling tasks or more generally.[35]

[34] There are results that suggest that schizophrenic individuals are less proficient at correcting experimentally-induced action errors under certain conditions than healthy individuals (e.g., Frith & Done, 1989; Stirling, Hellewell, & Ndlovu, 2001),

Another option is that the forward model prediction is *inaccurate*. The forward model prediction might fail, for example, to accurately project the spatiotemporal properties of the movement specified by the motor command. This, too, would help to explain the lack of sensory attenuation, since the inaccuracy would not yield a match at the comparator between the prediction and the sensory feedback, and so no sensory attenuation would take place.

While this explanation is tempting, there are two problems. First, once again, systematically inaccurate forward model predictions would likely yield sensorimotor deficits and schizophrenic individuals seem to exhibit intact sensorimotor control over their actions in general and within the self-tickling studies. Second, the successful explanation should be able to account not only for the lack of sensory attenuation in the self-tickling studies, but also for the action-monitoring results. The explanation in question fails when applied to these results, however, since what must be explained is that schizophrenic individuals are less accurate than controls at detecting mismatches between their movements and distorted visual feedback—they tend to judge that they match when they do not. The problem is that if the underlying issue were inaccurate forward model predictions, as this explanation would seem to hold, schizophrenic

---

but this is a highly specific and minor deficit compared to that which would be expected were they to fail to form forward model predictions altogether.

[35] One might be tempted here to appeal to the distinction that some theorists (e.g., Frith, 2005) draw between the so-called forward dynamic model, which is thought to predict the limb positions that would be consequent to the motor command, and the forward output model, which predicts the sensory consequences of the motor command. Based on this distinction, one might think that perhaps schizophrenic individuals are only impaired in the latter, and not the former, so that their sensorimotor control is still intact. But it is not at all obvious that a serious disruption in predicting the sensory consequences of an action, given the role this is thought to play in sensorimotor control, would not also lead to disruptions in such control. Moreover, independent motivation for positing both types of forward model is wanting.

individuals should be *less* likely than controls to report that distorted visual feedback corresponds to their movements, rather than *more* likely. Inaccurate forward model predictions would yield mismatches for schizophrenic individuals when compared with corresponding visual feedback from movements, whereas accurate forward model predictions would yield matches for healthy controls. We are looking for a unified explanation in terms of forward modeling to explain the results across paradigms, but we have not found one here.

Yet another option, which some theorists have recently favored (e.g., Synofzik, Thier, Leube, Schlotterbeck, & Lindner, 2010), is that the forward model outputs an accurate prediction, but that this prediction is imprecise. Unlike the previous proposal, this yields an attractive way to explain the action-monitoring results, since an imprecise yet accurate prediction should yield more frequent matches than predictions that are both accurate and precise, which healthy controls presumably generate, and this is just what we see. For example, suppose the prediction specifies that the sensory consequences of the movement will occur within a window of 100 – 200 ms after the movement onset, then any visual feedback with temporal properties within this range will register as a match, whereas a more precise prediction specifying that the sensory consequences occur within, say, 150ms will register mismatches for anything over or below this.

The problem with *this* proposal is that it cannot explain the *sensory attenuation* results, since an imprecise yet accurate prediction should still lead to sensory attenuation, as there would, in fact, be a match between the prediction and sensory feedback. Another issue with the present proposal is that it cannot explain delusions of control more generally. The issue, as others have pointed out (Synofzik et al., 2008; Voss et al., 2010),

is that if matches between forward model predictions and sensory feedback occur more frequently for schizophrenic individuals due to imprecise forward model predictions, and such matches underlie agentive awareness, then these individuals should not exhibit delusions of control, but rather excessive agentive awareness. They should be aware of themselves as performing certain actions when they are not. So while the proposal may help to explain the results of the action-monitoring studies, it cannot at the same time help explain the delusions of control that these studies themselves set out to illuminate in the first place. There is an irresolvable tension here.

A proponent of the CM approach might reply here that we ought to reexamine the assumption that any degree of match between the forward model prediction and sensory feedback is sufficient for a match to be registered. Rather, it may be that the match must exhibit a certain degree of precision in order for a match to be registered. After all, the CM is supposed to explain fine-grained motor control, where spatial ranges and temporal scales are programmed at a detailed level. This is also in line with the results of the perturbation condition in Blakemore, Wolpert, & Frith (1999), where small changes in spatiotemporal dimensions of the tactile stimulus canceled out the sensory attenuation effect. If so, then arguably matches resulting in sensory attenuation are not simply a function of whether or not the sensory effect is accurately predicted, but also a function of the degree of precision of the prediction. So even if imprecise forward model predictions accurately capture the range at which the sensory feedback occurs, they would, due to their imprecision, yield fewer matches than their precise counterparts, and so one would not expect to see as much sensory attenuation for schizophrenic individuals as for healthy controls.

Suppose that we grant this revised assumption about what is required for a match and thereby sensory attenuation to take place. Now the proposal, once again, faces a problem explaining the action-monitoring results. For it would follow, if this proposal were correct, that schizophrenic individuals would register more mismatches than healthy controls between their forward model predictions and visual feedback. For example, suppose again that the forward model prediction specifies that the upcoming bodily movement will occur within 100 – 200 ms. Given the precision requirement that we have temporarily adopted to accommodate the sensory attenuation results, visual feedback with temporal properties in that range will nonetheless fail to register as a match—but this is opposite to the result that we see of overattribution of visual feedback.

Where does this leave us? It would seem that we are out of options for an explanation in terms of forward modeling that can both explain the results of the relevant studies and explain the behavior of schizophrenic individuals outside of these experimental settings.

If breakdowns in forward modeling cannot explain the performance of schizophrenic individuals on action-attribution and sensory attenuation studies, what *does* explain them? As mentioned in Chapter 2, it has been well-established in the empirical literature that schizophrenic individuals exhibit impairments in attention and sensory processing. So an attractive alternative explanation is that these impairments affect their performance on the relevant tasks as well.

There is ample evidence that schizophrenic individuals are impaired with respect to visual processing. For example, schizophrenic individuals reportedly have difficulty with figure-ground perception, i.e., telling objects apart from the background against

which they are presented (Malaspina et al., 2004). It has also been documented that they show deficits in perceptual closure, i.e., identifying an object that is partially occluded or fragmented (Doniger, Silipo, Rabinowicz, Snodgrass, & Javitt, 2001).

In addition, mounting evidence suggests that schizophrenic individuals have deficits in attentional processing. More specifically, several studies suggest that schizophrenics are impaired in selectively attending to relevant stimuli while ignoring irrelevant distractors (selective attention) and in maintaining focus throughout tasks (sustained attention). For example, one study found that they were less able to control top-down attention in order to identify a target in a visual search task relative to controls (Gold, Fuller, Robinson, Braun, & Luck, 2007). And another study found that they were slower and less accurate than controls on a visual selective attention task in which they were to respond to the presence of target stimuli (Carter et al., 2010; Fioravanti, Carlone, Vitale, Cinti, & Clare, 2005).

These visual processing and attentional impairments could help explain the performance of schizophrenics on action-attribution studies. If, as I have argued, individuals use the strategy of comparing the visual feedback they receive with proprioceptive feedback from the movement, then visual and attentional impairments, would result in schizophrenic individuals being less accurate than controls at detecting mismatches between visual feedback and proprioception in the action-monitoring task.

The sensory attenuation results could also be explained by appeal to perceptual or attentional impairments in schizophrenic individuals. Once again, if the alternative explanation I proposed in terms of cognitive expectations and greater attentional resources being devoted to externally- vs. self-generated tactile stimuli, then deficits in

attentional processing could affect their performance on the self-tickling tasks as well. . Perhaps they are unable allocate their attentional resources in a way that is proportional to the salience of the stimuli in question, and so the externally-generated and self-generated stimuli are perceived on a par with one another.

Of course, much remains to be explained regarding the nature and extent of the perceptual and attentional deficits that schizophrenic individuals are thought to possess. And an explanation of their performance on these tasks by appeal to these deficits must be further explored and evaluated. One issue in particular is that these impairments seem to be general impairments that schizophrenic individuals possess, whereas some of the results we discuss seem to implicate specific deficits for schizophrenic individuals with delusions of control. It will be important to understand in future work, therefore, to what extent attentional and perceptual deficits are correlated with the types of specific symptoms schizophrenic individuals exhibit. The point here is simply that there are credible alternative explanations for their performance on the tasks that we have discussed that make no appeal to forward modeling, and explanations in terms of forward modeling do not, upon closer examination, hold much promise in giving a unified explanation of the experimental data, as well as delusions of control more generally.

§3.4 Divorcing the Intentional Binding Effect From Forward Modeling

Forward modeling has also been appealed to in order to explain the intentional binding effect. But there is evidence that strongly suggests that an explanation in such terms is misplaced here as well. The problem is that proponents of the CM approach have often assumed that intentional binding is linked exclusively to agency. There is evidence, however, that rather than having to do with action *per se*, intentional binding is instead a

type of *temporal* binding that occurs when one is aware of a causal relationship holding between two events more generally (see Synofzik, Vosgerau, & Lindner, 2009, p. 1066) (see also Synofzik, Vosgerau, & Lindner 2009, 1066). In a study by Wöhlschlager, Engbert, & Haggard (2003), participants were asked to report the onset time of a lever press in three separate conditions. In the "self" condition, participants pressed the lever spontaneously—they were asked not to plan out the timing of their actions. In the "other" condition, participants observed the experimenter pressing the lever. In a third condition, the "machine" condition, they observed a machine pressing the lever. In the relevant experiment, the lever presses were followed after 250 ms by an auditory tone. The results showed a binding effect, i.e., a delay in the perceived onset of the action shifted toward the subsequent tone, in *both* the "self" and "other" conditions. This suggests that binding occurs not merely in cases where one is oneself performing an action, but in cases in which one is observing another agent performing an action. But since forward model predictions do not underlie the observation of action, an explanation in terms of forward modeling will not do here.

A second, more recent, study by Buehner (2012) converges on the same conclusion. In the study, participants were asked to predict when a target light would flash, and given different predictive cues to go on in each condition. In the baseline condition, the flash was preceded by an LED signal, in the "self-causal" condition, participants pressed a button to generate the flash, and in a "machine-causal" condition, the button was pressed by a machine, which generated the flash. The target flash followed the predictive cue in fixed intervals of 500 ms, 900 ms, or 1300 ms.

The results showed binding between the predictive cue and the target light in both the self-causal and machine-causal conditions. Once again, this suggests that the binding effect is not underwritten by forward model predictions posited in the CM, since it does not depend on being the agent of the event that gets temporally bound to its effect. Rather, it appears to be the result of representing a causal relationship between two events, more generally.

Not only do the above results undermine the connection between intentional binding and forward modeling, they also call for a rejection of the claim that intentional binding is an implicit measure of agentive awareness, or is somehow responsible for generating the agentive awareness, as many have suggested. Indeed, even if intentional binding were shown to occur exclusively when one is the agent of the action being bound to its effect, these stronger claims would remain unsupported, since they are claims about one's *awareness* of being an agent, and *this* has not been shown to be related to intentional binding.

In this section I have challenged the assumption that many theorists make, either implicitly or explicitly, that sensory attenuation, action monitoring, and intentional binding somehow rely on forward modeling. But this is a crucial premise in the case for the CM approach to agentive awareness. Without this premise, the CM approached is stripped of its main support.

§4 The Case Against the CM Approach to Agentive Awareness: Dissociating Agentive Awareness and Sensorimotor Control

The CM approach to agentive awareness relies on there being a tight connection between sensorimotor control and our awareness of ourselves as acting. More

specifically, such control is thought to be both necessary and sufficient for agentive awareness. After all, when and only when such control is operative is there a match between the forward model prediction and sensory feedback from the bodily movement. And it is this match that is itself thought to be necessary and sufficient for agentive awareness, on the CM approach. In this section, I challenge the CM approach to agentive awareness by arguing that sensorimotor control and agentive awareness dissociate.

§4.1 Sensorimotor Control Without Agentive Awareness

I start by presenting some everyday examples of sensorimotor control without agentive awareness. Perhaps most striking and familiar are cases of "action slips", one of the most common of which is a *capture error* (see Reason, 1990). Capture errors take place when one performs a habitual action that interrupts one's intended action. William James (1890/2007) famously described the following capture error:

> … very absent-minded persons in going to their bedroom to dress for dinner have been known to take off one garment after another and finally to get into bed, merely because that was the habitual issue of the first few movements when performed at a later hour. (p. 115)

During such an action slip, and others like it, one does not have agentive awareness over what one is doing. If one did, one would inhibit the actions in question, given that they conflict with one's intentions. However, at odds with the predictions of the CM approach, these actions exhibit sensorimotor control—they are perfectly well-executed in smoothness, coordination, and timing, and in terms of their interactions with environmental objects. It is only relative to one's intentions, that is, at the level of intentional control, that they are aptly characterized as "slips". And as Bayne (2011b)

acknowledges: “Crucial to the comparator approach is the notion that agentive awareness can be generated by mechanisms that need not—and typically will not—have access to fully-fledged intentions” (p. 495). But this does not seem to deliver the correct result here.

One may object to this type of example on the grounds that, once one realizes what one is doing, one does not *disown* the actions in question, and this suggests that one does, in fact, have agentive awareness at some level during the action. But attributing an action to oneself after the fact is not sufficient for agentive awareness. Agentive awareness requires a subjectively unmediated awareness of what one is presently doing, and in action slips, one comes to be aware of oneself as acting by way of conscious observation. One suddenly sees, for example, that one is pouring orange juice instead of milk into one’s cereal bowl. Or one “finds oneself” having taken one’s usual route home instead of taking the appropriate turn to stop at the grocery store. One *discovers*, as it were, what one is up to as an agent. This is at odds with how one is typically aware, in a subjectively unmediated way, and in the way characteristic of agentive awareness, of one’s present actions. It is for this reason that I characterize these cases as involving an absence of agentive awareness.

Moreover, even if a post-hoc sense of authorship were sufficient to establish that agentive awareness is present in these cases, it is not clear that one always *does* claim authorship over these actions after the fact. Consider the following anecdote, related by Marcel (2003), concerning a friend, Jonathan Miller:

> As a young man, he came across a crowd who, he learned, were waiting for the Queen to pass in a car. Although this was one of the last things Jonathan would

> consider doing, he could not easily get past. When the car passed, the crowed cheered and waved. […] Jonathan says that to his surprise and horror, given his self-image, "I found myself raising my arm and starting to cheer," which he quickly suppressed. One thing that embarrassed him was that his behaviour was something which seemed under his voluntary control and could have been inhibited even earlier had he been aware of it before. Yet he says that he did not feel that the actions (raising his arm and cheering) were 'his'. He supposes that they were induced by a kind of contagious imitation. (p. 75)

It is reasonable to view Jonathan's arm-raising and cheering as being under the purview of sensorimotor control, given that the likely triggers were environmental, the behavior was nonconscious, and at odds with his intentions and desires. And, like the other cases of action slips we have discussed, Jonathan reports that he "found himself" raising his arm—he became aware of what he was doing by conscious observation. And yet, in line with what I have been urging, he did not feel, even after the fact, that he was the author of this action in some important respect. Of course, this is only an anecdote, and more robust evidence is required before the case is settled. But it is worth bearing in mind that the proponent of the CM approach has no more than anecdotal evidence supporting the fails to get the upper hand here.

In the clinical literature, too, we find cases where there is no agentive awareness despite intact sensorimotor control. Here we may appeal once again to anarchic hand syndrome. This condition serves as a particularly good example because one of its identifiable markers is that the anarchic hand is "disproportionately reactive to external environmental stimuli" (Biran, Giovannetti, Buxbaum, & Chatterje, 2006, p. 563), and

this is one of the key markers of sensorimotor control. Recall the case of JC from Chapter 2, who acquired anarchic hand syndrome at the age of 56 after suffering a stroke. JC's anarchic limb executes well-coordinated movements at the level of sensorimotor control, such as reaching and grasping for objects, turning taps on and off, pressing keys, and so on. But JC is not aware of the actions as his, reporting that the hand "[h]as a mind of its own" and "[w]ants to be the boss" (Biran et al., 2006, p. 567).

It is also worth noting that JC may only be aware of his anarchic actions when he is consciously viewing them. Consider the following incident, in which JC, busy with a different task and therefore not observing the anarchic limb, was unaware of the limb's movements:

> For example, in one of the testing sessions, [JC] was asked to turn pages of a magazine with his left hand. As he did this (without any difficulty), the examiner lightly touched his right fingers with a pen. The right hand reached towards and persisted in following the pen continuously as it was slowly moved away from the hand […]. This reaching continued until the limb was a foot above the table. JC was unaware of his arm moving on that side. (Biran, Giovannetti, Buxbaum, & Chatterje, 2006, p. 567)

More evidence is needed here, but it may very well be that unless the actions of the anarchic limb are consciously observed, one has no conscious awareness of them, and certainly no agentive awareness of them. But once again this would seem to be at odds with the predictions of the CM approach to agentive awareness.

The proponent of the CM approach could point out that the CM also posits a comparison between the agent's intended state and sensory feedback, and that a

mismatch *here* could be responsible for the lack of agentive awareness in the case of anarchic hand syndrome. But going this route just assimilates the CM approach to a high-level account, since what is in effect being appealed to is a mismatch between the agent's intention and what she perceives herself to be doing. The CM approach as a purely *low-level* approach does not seem to have the resources to explain why, in the case of anarchic hand syndrome, the individual does not have agentive awareness—and this is the target here.

§4.2 Agentive Awareness Without Sensorimotor Control

I have now argued, by way of some concrete cases from everyday and clinical contexts, for one half of the double dissociation between agentive awareness and sensorimotor control. I now turn to arguing for the second half, which requires cases of agentive awareness despite disruptions in sensorimotor control.

When an error occurs in the execution of an action, and is subsequently corrected, this entails, on the CM, that a mismatch occurs between the forward model prediction and sensory feedback from the movement. If there were no mismatch, there would be no error correction. So we can look at cases of error correction to evaluate the claim that when a mismatch occurs, there is no agentive awareness. Looking at these cases, however, suggests that even in cases of mismatch, agentive awareness persists.

Consider, first, everyday cases in which an action is performed in a way that reflects impaired sensorimotor control, but where there is nonetheless no accompanying interruption in agentive awareness. Suppose you are asked to write your name with your non-dominant hand. If you are like most people, you will experience significant difficulty in controlling the fine-grained movements of your hand so that your pen moves

appropriately across the page. You will likely correct many small errors as you go along. Despite this interference with the sensorimotor control of your action, you do not, as a result, lack agentive awareness of it. You are aware, in a subjectively unmediated way, that you are writing your name, and you naturally attribute the action to yourself.

Experimental work also provides evidence for the claim that agentive awareness is sometimes present despite impaired sensorimotor control. I will consider three sources of such evidence.

First, in a popular experimental paradigm known as the *double-step paradigm*, participants are asked to point to targets that suddenly change position midway through the participants' pointing movements. Castiello, Paulignan, and Jeannerod (1991) asked participants to sit at one side of a table, on the other side of which were three vertical rods that may or may not light up. Participants were told that if they see a rod light up, which occurred on 20% of the trials, they should reach and grasp it. But, on these trials, as soon as the participant started reaching for the lit target rod, it switched off and the one beside it lit up. In order to follow the experimenters' instructions, participants had to adjust the trajectory of their reaching movement so that they successfully grasped the new target. Participants had no trouble performing these online adjustments of their movements. Indeed, their response to the target shift occurred on average about 110 ms after it happened. There was an additional twist to the study, however. Participants were also asked to shout out 'TAN' as soon as they actually *saw* the target shift. They would typically only do this at about 420 ms, or 300ms *after* they had already corrected their movements. So they were not aware of the shift at the time that they corrected their movements. Nor were they aware of having already made the required adjustments. The

upshot is that sometimes action errors are both registered and corrected and yet there is no effect on agentive awareness.

In another study using a paradigm similar to the action-monitoring paradigm discussed earlier, Fourneret and Jeannerod (1998) asked participants to draw a straight line with a stylus on a tablet. They could not directly see the movements they were performing, but instead faced a mirror that purportedly showed them the line they were drawing. Sometimes, though, the experimenters altered the visual feedback, so that it appeared that the line the participants were drawing deviated either to the left or to the right. This meant that the participants had to adjust their own movements in a corresponding way in order to successfully draw straight lines—which they did accordingly. Strikingly, though, the participants' verbal reports did not reflect any conscious awareness of having altered the trajectory of their movements in light of the misleading visual feedback. Nor were there any disruptions in agentive awareness, despite the disruptions that must have been present at the level of sensorimotor control in order for the compensatory adjustments to take place.

Third, in an elegant study by Logan and Crump (2010), skilled typists were asked to type single words presented to them on a computer screen. Their typed responses appeared below the word they were to type on the screen. On some trials, the experimenters inserted errors in correct responses that typists made, and on other trials, they corrected errors that typists made. So, for example, if the typists were asked to type the word 'beautiful', the experimenters might insert the error 'beauytiful'. The results were that the typists took credit for the corrected errors that they saw on the screen and accepted blame for the inserted errors. That is, they judged that they had typed the word

correctly on trials in which the experimenter had corrected their error, and that they had typed the word incorrectly on trials in which the experimenter inserted an error. On the face of it, this suggests that they did not register the errors they made, but in fact their typing rate revealed otherwise. More specifically, their typing rate slowed down after they made errors that were subsequently corrected by the experimenters on the screen, but not after the errors inserted on the screen by the experimenters. Post-error slowing is a common phenomenon in sensorimotor control. This suggests that an error was indeed registered at that level, and yet, once again, this had no effect on agentive awareness.

Finally, we may turn once more to the clinical literature for evidence of a dissociation between agentive awareness and sensorimotor control. Recall anosognosia for hemiplegia, which involves the individual's denial of contralesional paralysis. Patients with this disorder report being aware of themselves as performing certain actions with their paralyzed limb, seemingly oblivious to the complete absence of movement. On the face of it, this would certainly seem to suggest agentive awareness in the absence of sensorimotor control. Where there is no movement, there is, of course, no such control.

The foregoing discussion supports a double dissociation between sensorimotor control and agentive awareness. This poses a serious problem for low-level approaches to agentive awareness, and specifically the CM approach, since these approaches require that agentive awareness be anchored in sensorimotor control. I proceed in the next section to consider some of the replies the proponent of the CM approach may offer here.

§4.3 Objections and Replies

There are several replies available to the CM theorist when confronted with the double dissociation that I have been arguing for between agentive awareness and

sensorimotor control. Regarding the claim that sensorimotor control is sometimes present while agentive awareness is absent, proponents of the CM approach to agentive awareness may reply that it is not enough for sensorimotor control to be exercised, in addition the agent must be *aware* of having such control in order for agentive awareness to result. Indeed, Frith (2005) himself gestures towards this reply in connection with schizophrenic individuals with delusions of control, when he writes that, "… the problem is not with motor control, but with *awareness* of motor control" (p. 766, emphasis mine).

The revised proposal just restates the problem we started out with, though. The challenge for any theory of agentive awareness is to specify the states in virtue of which we are agentively aware, and how they arise. So to say simply that we are agentively aware in virtue of being aware of our sensorimotor control is not helpful here, since it does not specify the states or mechanisms underlying *this* kind of awareness. Indeed, it is difficult to see how this type of view would still honor the commitments of the CM approach, since the states in virtue of which one is aware of having sensorimotor control would presumably not be states posited within the CM itself, but states existing outside that system. It seems we would need to look to high-level states for an account of agentive awareness along those lines, so this would not help the low-level approach.

Another strategy for rejecting the claim that sensorimotor control is sometimes present without agentive awareness is to argue that one *does* have agentive awareness in these cases, but the states in virtue of which one is so aware are themselves non-conscious, and so the agent is unable to report on them. Indeed, the states of the comparator model are standardly characterized as "sub-personal" (Synofzik et al., 2008a; Hohwy & Frith, 2004; Bayne, 2008), a term owing to Dennett (1969), which refers to

states of a person's subsystems, e.g., states of the digestive system, rather than states of the person, e.g., beliefs, desires, intentions. And, arguably, and as many theorists hold, only our personal level states are ever conscious. For example, Bermudez (2010) writes, "… subpersonal mechanisms operate completely below the threshold of consciousness (there is no such thing as subpersonal awareness" (p. 587). But if all the states involved in the comparator model are subpersonal states, as many claim, then there is no reason to think that that they are *ever* conscious—so how can they ever be the states in virtue of which we are agentively aware?

At this point, the proponent of the CM approach can either abandon the claim that the states of the CM are sub-personal, or claim that sub-personal state are *sometimes* conscious—and it is at these times that we have agentive awareness. They tend to go with the latter option. For example, Synofzik et al. (2008), in discussing the CM approach to agentive awareness, hold up the following as a virtue of the view: "This registration [of a match between forward model prediction and sensory feedback] is not necessarily conscious or explicit, but in fact occurs mostly in an implicit, pre-conscious way" (p. 222). Their use of "mostly" allows that these states are conscious at least some of the time. But, more specifically, Blakemore, Wolpert, and Frith (2002) write: "… we seem to be unaware of the results of the comparison between the predicted and intended outcome of motor commands, and the comparison between the predicted and actual sensory feedback, as long as the desired state is successfully achieved" (p. 237). And Pacherie (2008) writes, "[i]t is only when the discrepancy between predicted and actual sensory consequences of the movements becomes too large to be automatically corrected that it becomes accessible to consciousness" (p. 210). The idea here seems to be that the

output of the comparison is conscious only in cases where there is a large mismatch between the predicted and actual sensory consequences of an action.

While this gives an indication of when the states of the CM are conscious, it leads to the problematic result that the CM approach only explains how we come to have the phenomenology that something has gone *wrong* with our actions. It does not provide an explanation for agentive awareness since, according even to proponents of the view, when everything is going smoothly, the states involved in the relevant comparison are not conscious. It does not look like the CM approach has a promising way around cases in which sensorimotor control is intact, and yet there is no agentive awareness.

Moving on, the proponent of the CM approach to agentive awareness may also take issue with the claim that, in some cases, sensorimotor control is impaired while agentive awareness is undisturbed. A possible reply here is that the sensorimotor impairment in question is not large enough for there to be a disruption in agentive awareness, so there is still *enough* sensorimotor control to underpin agentive awareness in some of the cases I offered as evidence for this dissociation. As Synofzik, Vosgerau, & Newen (2008) urge, "[n]ot any mismatch between predicted and actual state should lead to an evaluation of the event as being externally produced, but only mismatches beyond a certain sensitivity range" (p. 223). (The earlier quotation from Pacherie also echoes this claim.)

This reply runs the danger of creating *ad hoc* conditions for the theory. For in any case in which there appears to be a disruption in sensorimotor control, but no disruption in agentive awareness, the proponent of the CM approach is free to claim that the

mismatch underlying the disruption is not sufficiently large. But what counts as sufficiently large here? The theory makes no predictions.

A second problem with this reply is that the looser the match, and the greater the mismatch, that is required in order for agentive awareness to arise, the more the CM approach starts to look like a high-level approach. Part of what it is to give an account in terms of the CM and sensorimotor control is to attribute to states that represent fine-grained features of action the role of determining agentive awareness. But if agentive awareness turns out not to be a function of the detailed content of such states, then it is unclear why one should insist that it is the CM that is responsible, rather than a higher-level mechanism. The CM approach appears to be stripped of its motivation if one takes this reply on board.

§5 Conclusion

The CM may be a successful model for explaining sensorimotor control. But I have argued in this chapter that it faces grave difficulties when it comes to explaining agentive awareness. The empirical results that are so often cited in its favor do not, upon closer examination, support a connection between the CM and agentive awareness. And, moreover, there is a robust double dissociation between sensorimotor control, which is thought to be modeled by the CM, and agentive awareness. As such, I conclude that we must look elsewhere, to high-level accounts, for a theory of agentive awareness. I turn to this task in the next two chapters.

CHAPTER 4: THE MATCHING APPROACH TO AGENTIVE AWARENESS

§1 Introduction

In Chapter 3, I argued against low-level accounts of agentive awareness, taking the CM version of such an approach as my target, as it is by far the most dominant. Since both low-level and integrated approaches to agentive awareness take the states and processes involved in the CM to make important contributions, if the arguments of Chapter 3 are successful, then this clears the way for a high-level account of agentive awareness cast in terms of mental states like intentions, beliefs, and perceptions.

Common to all the main accounts of agentive awareness is the claim that agentive awareness arises out of a *match* between a goal state (e.g., an intention, motor command, or forward model prediction), and an action-monitoring state (e.g., a sensory or perceptual state) (e.g., Pacherie, 2008; Wegner, 2002). In this chapter, I will challenge this popular claim as well, arguing that if a high-level account is indeed correct, it is not a matching version that we seek.

I will treat as my focus the most popular high-level matching account of agentive awareness, which is that defended by the psychologist Daniel Wegner (2002) in his book *The Illusion of Conscious Will*. Though I aim my discussion at this account, I take it to generalize to other versions of the high-level matching approach. Wegner's primary aim in the book, and elsewhere (Wegner, 2004; Wegner & Wheatley, 1999) is to establish that agentive awareness, what he labels the *experience of conscious will*, is in some significant way misleading. This is an important claim worthy of attention in and of itself, but it will not be my focus here. I will be interested, rather, in the account Wegner provides of how agentive awareness, whether illusory or not, comes about.

I begin in the next section by describing Wegner's model for the psychological mechanism underlying agentive awareness. In section 3, I evaluate the case in favor of Wegner's view and argue that it faces several difficulties. In section 4, I offer a simple challenge against Wegner's view and all matching models to boot: Agentive awareness sometimes arises in the absence of any bodily movement, so no match is required. I conclude by considering and replying to objections to this challenge.

§2 Wegner's Theory of Apparent Mental Causation

To begin with, it is worth stepping back to consider the question of why a matching account of agentive awareness is attractive to so many in the first place. Perhaps the most inviting feature of such an account is that it offers a straightforward explanation of how it is that we come to distinguish between bodily movements of ours that are actions and those that are not. A correspondence between what one intends to do and the events, whether bodily or otherwise, that one perceives following one's intention is good evidence that a causal relationship holds between them. And on the control theory of action that I favor, which lines up nicely with our commonsense view of action, actions are, roughly, events that are suitably caused by our intentions. So it makes sense to think that evidence that one's intention is the causal source of corresponding events gives rise to awareness of those events as one's actions. On this view, then, agentive awareness results from simple causal reasoning about the relationship between our intentions and the events we perceive.

With this in mind, let us take a closer look at Wegner's own version of the matching model (Wegner, 2002, 2004; Wegner & Wheatley, 1999). On Wegner's

particular account, which he calls the *theory of apparent mental causation*, one is aware of oneself as the agent of an event when one infers that one's intention causes it.[36] Wegner (2004) views these inferences as special instances of causal reasoning more generally. As he himself puts it,

> [t]he person experiencing will, in this view, is in the same position as someone perceiving causation as one billiard ball strikes another. As we learned from Hume, causation in bowling, billiards, and other games is inferred from the constant conjunction of ball movements. It makes sense, then, that will – an experience of one's own causal influence – is inferred from the conjunction of events that lead to action. (p. 654)

The causal inference underlying agentive awareness, on Wegner's view, is sensitive to three conditions: the priority condition, the exclusivity condition, and the consistency condition. The priority condition requires that the intention that one infers is the cause of the action "appears in consciousness just before [the] action" (Wegner, 2002, p.66). The exclusivity condition requires that the relevant intention is "not accompanied by conspicuous alternative causes of the action" (Wegner, 2002, p.66). Finally, the consistency condition requires that the intention in question is "consistent with the action" (Wegner, 2002, p.66). The model is straightforward. Suppose I form an intention

[36] This leaves it open as to what kind of mental state possesses the phenomenal properties associated with the experience of conscious will. One option is that it is the mental state embodying the conclusion of the inference, i.e., a belief or thought that one's intention causes one's action. Another option is that it is an affective state that is triggered by the inference. Indeed, Wegner (2004) suggests as much when he refers to the experience of conscious will as an "authorship emotion" (p. 667). I will stay neutral as to which interpretation is to be preferred, as it will make no difference to the particular arguments I present in this chapter.

to raise my arm. On Wegner's picture, if I am aware that I have an intention to raise my arm, and then, immediately after, I perceive that my arm is going up, and I am aware of no salient alternative cause of my arm's going up, then I will likely infer that my arm's going up was caused by my intention to raise my arm, and I will be aware of *myself* as raising my arm—I will be aware of my arm's going up as my action.

In support of the priority condition, Wegner (2002) appeals to work by the Belgian psychologist Albert Michotte (1946). Michotte was interested in understanding the nature of our causal reasoning processes in general. He found that when people view an apparent object strike another object, and the second object moves immediately and in the same direction as the first, people tend to judge that the movement of the second object is caused by the first object. Perhaps unsurprisingly, such judgments are typically blocked, he found, if the second object starts to move too long after the first object makes contact with it, or if it moves before the first object makes contact with it. Inspired by Michotte's work, Wegner concludes that "to be perceived as truly a cause, an event can't start too soon or too late; it has to occur just before the effect" (p. 71). Applied to agentive awareness, this commonsense constraint on causal reasoning yields the priority condition—an intention must occur just before an event in order for one to infer that it is the (proximal) cause of that event.

As for the second condition, the exclusivity condition, this once again lines up with a commonsense picture of causation. We tend to attribute a single proximal cause to events, so if there are competing events that one may view as causes of some event, the chances that any one particular event will be identified as the cause are reduced.

Wegner's thought is that this principle applies no less to mental events like intentions than it does to other events in the world to which we attribute causal powers.

When it comes to the final condition, the consistency condition, Wegner (2002) avers that people more readily attribute causal relationships between two events when "they can appreciate some meaningful connection between a candidate causal event and a particular effect" (p. 79). As he writes elsewhere, expanding further as this applies to intentions and their effects:

> [a] thought that is perceived to cause an act is often the name of the act or an image of its stimulus, execution, or consequence […]. Consistency of thought and act depends on a cognitive process whereby the thoughts occurring prior to the act are compared with the act as subsequently perceived. When people do what they think they were going to do, there exists consistency between thought and act, and the experience of will is enhanced. When they think of one thing and do another—and this inconsistency is observable to them—their action does not feel as willful. (Wegner & Wheatley, 1999, p. 485)

So, once again, we can view Wegner's model as a high-level version of the comparator model, in which an intention and perception of an action are compared, and a successful match leads to an inference to the effect that one's intention has caused one's action.

In the next section, I evaluate the evidence that has been presented in favor of Wegner's matching model, and, indeed, in favor of any similar account that takes a match between an intention and an action to be necessary for agentive awareness.

§3 Evaluating the Empirical Evidence for a Wegner-Style Matching Account

In this section, I examine the experimental work that has been widely cited and celebrated as offering support for a Wegner-style matching model of agentive awareness. I do think, as I will explain in Chapter 5, that there is some merit to this model for explaining some aspects of how we come to recognize our own agency, though it is unlikely, as I will also argue in this chapter, that it captures the mechanisms underlying agentive awareness. But, at any rate, the success of the empirical case in its favor has been greatly exaggerated in the literature. My aim in this section is to show that this is so, with the hope of motivating more careful work in this area in the years to come.

§3.1 'I Spy' Study

A study that is commonly cited as support for a Wegner-style matching model is the 'I Spy' study conducted by Wegner and Wheatley (1999). In one experiment in this study, participants were paired with confederates and seated across from each other with a square board in between them that was mounted on top of a computer mouse. Both the participant and the confederate were asked to place their fingertips on the board so as to move the mouse together, simulating a 'ouija board' set up. They were asked to move the mouse in "slow sweeping circles" (p. 487), which would move a cursor on a computer screen that was clearly visible to both of them. On the screen was a photo from the book *I Spy* showing a number of small objects (e.g., car, plastic dinosaur, swan).

After each 30-second interval, there was a 10-second interval during which the participants and confederate were supposed to stop the cursor on one of the objects. During this "stop" interval, the participants would hear music and a word over their headphones naming an object on the screen. They were told that they would hear different words than those heard by the confederate, and that the words were meant to

serve as "mild distractions" (p. 488). In fact, the confederate heard instructions to move to a particular object on the screen at a particular time, which matched the object that the participants heard over their headphones. The timing of the confederate's stop was such that the participant would hear the word corresponding to the object the confederate stopped on 30-seconds, 5-seconds, or 1-second *before*, or 1-second *after* the confederate stopped on the object. On the rest of the trials, the confederate allowed the participant to make the stops and did not interfere. In these cases, the participants heard a word 2-seconds into the 10-second "stop" interval. The word corresponded to an object on the screen for roughly half of these "unforced stop" trials.

After each trial, participants would rate the extent to which they intended to make the stop. They did so by placing marks on a scale that had one endpoint indicating 'I allowed the stop to happen' and another endpoint indicating 'I intended to make the stop'. The participants' marks on the scale were then converted to percentages.

On average, the participants rated the forced stops, that is, those made by the confederate, at around 52% on the scale. Moreover, the degree to which they rated the stops as intended increased the closer the priming word occurred to the stop, with the average rating at around 44% when it occurred 30-seconds before the stop, and climbing up to between 55 – 60% as it approached 5-seconds and 1-second before, then dropping down again to around 45% when it occurred 1-second after the stop.

From these results, Wegner and Wheatley (1999) conclude that, "… the experience of will can be created by the manipulation of thought and action in accordance with the principle of priority" (p. 489). And many others have followed them in this

conclusion. For example, in discussing this experiment, Levy and Bayne (2004), write that:

> Wegner shows that if participants are primed to think about a particular image on the screen, *they are likely to experience themselves as moving the cursor to that image*, and hence, believe that they moved the cursor to that image, even when the experimenter rather than they themselves was responsible for moving it. (p. 463, emphasis mine)

And, more recently, Moore et al. (2009) write that the experimenters managed to "induce a false sense of agency for movements that participants had not in fact performed" (1056). (For similar endorsements of this conclusion in the literature, see also Blakemore & Frith, 2003, p. 220; Roskies, 2010, p. 126).

There are several problems here. First, as others have noted (e.g., Carruthers, 2012), what the participants were explicitly asked to judge is the degree to which they *intended* to make the stops, not the degree to which they were agentively aware of making them. And, supposing we take these judgments at face value, intending to do something is not the same as being agentively aware of doing it, though it is perhaps necessary for it. When one performs an action on "autopilot", for example, such as locking one's front door when going out, one is frequently not agentively aware of oneself as doing it at the time. And yet one might judge after the fact that one intended to do it. So the conclusion drawn by many to the effect that agentive awareness was induced in this study on the basis of the participants' ratings of how much they intended to make the forced stops is not warranted.

In addition, it is not clear that participants really did judge that they intended the forced stops. On average, they rated the degree to which they intended the forced stops as barely more than *halfway* up the scale. But if, on average, participants really did infer that they intended the forced stops, then one would expect them to mark the scale at or near 100%, on the side of the line that is explicitly labeled 'I intended to make the stop.' So not only can the results not be used in support of the claim that the participants had illusory agentive awareness for the forced stops, it is doubtful that they can even support the claim that participants inferred that they intended the stops.

Even if participants did sometimes infer that they intended the forced stops, it is not clear that Wegner and Wheatley (1999) ruled out that they did, in fact, intend to stop on the objects named in their headphones on these trials. If so, it is not clear that any false beliefs regarding their intentions were successfully induced by the experiment. The experimenters report doing an analysis on the *unforced* stop trials in order to determine whether the participants "might naturally stop on" the objects they were primed with in these cases, which might suggest that they were also intending to do so in the forced stop trials. They report finding no significant difference, on these unforced trials, in the mean distance between the final location of the cursor and the object for unforced stop trials on which an object on the screen was named, and unforced stop trials on which no object was named, concluding that "the forced stops created by the confederate were thus not likely to have been abetted by movement originated by the participant" (p. 489). But it is unclear what was used as the "object" in measuring the distance from cursor to object for trials in which there was no object named on the screen. So we cannot conclude that the

experimenters ruled out that participants actually did sometimes intend to move towards the object named in the forced stop trials.[37]

Even if we bracket the above points, and we grant that participants were agentively aware of making the stops in forced trials, this experiment still does not lend support to Wegner-style high-level matching models that we are presently considering. Wegner is not careful to distinguish between thoughts and intentions, nor does he distinguish between thoughts about doing something and thoughts unrelated to action. In the 'I Spy' study, according to the experimenters themselves, the spoken-word primes did not cause the participants to intend to stop on the objects named—though I still maintain my concerns about this conclusion from earlier—they merely, if anything, caused them to have a thought about the relevant object, and to be subsequently aware of themselves as intending to make the relevant stop. So this study cannot establish that agentive awareness requires a match between one's intention and one's action, which is what must be shown here.

Another issue is that participants in the 'I Spy' study could have at least sometimes arrived at their ratings by way of post-hoc conscious theorizing, rather than recalling that they actually intended or were aware of themselves as the agents of the actions in question. But then the most that this study shows—setting aside momentarily the significant concerns raised above—is that in ambiguous agency contexts, participants can sometimes be manipulated into reporting that they (somewhat) intended an action when they in fact did not. This does not alone entail anything about agentive awareness,

[37] Indeed, in post-experiment interviews, some participants reported sometimes "searching for" (Wegner & Wheatley, 1999, p. 489) the items they heard named over their headphones, which suggests that they sometimes did intend to stop on them. It is not clear why they were not asked this question directly in the interview.

which is the *subjectively unmediated* awareness of oneself as performing an action—it is not the result of conscious theorizing about oneself as an agent, as may be the case here.

A final problem with the 'I Spy' study is that Wegner's own exclusivity condition is not met, so by his own account, agentive awareness should not be generated. There is another agent that is a competing cause for the action in question. Indeed, the participants seemed sensitive to the ambiguity of the situation, since, as mentioned, even on trials when they were free to stop on whichever object they liked, without interference from the confederate, they rated the intended nature of their stop at 56.09%. One explanation of these low ratings is that participants felt that they *neither* allowed the stop to happen, *nor* intended it to happen. This might be the case if, for example, they had no intention of stopping the cursor on a particular item, but felt that they still contributed to the stop by moving the cursor jointly with the experimenter. Their mid-range ratings on the scale from 'I allowed the stop to happen' to 'I intended to make the stop' are just what would be expected on the basis of this interpretation.

In sum, Wegner's "I Spy" study cannot be appealed to as evidence for a high-level matching model of agentive awareness.

§3.2 The 'Helping Hands' Study

Another purported source of evidence for a Wegner-style matching model comes from the 'helping hands' study conducted by Wegner, Sparrow, and Winerman (2004). In this study, participants watched themselves in the mirror while another subject—a "hand helper"—stood behind them and extended their hands forward on either side of the participant. From the participant's point of view, another person's arms and hands were located where their own arms and hands would normally be.

The "hand helpers" heard a sequence of instructions over headphones, such as "wave hello with your right hand," and "give the 'OK' sign with both hands." Participants were told that they would sometimes hear instructions over their headphones, and that if they did hear instructions, they would sometimes relate to the actions of the hand helper. In one condition ("preview condition"), participants heard the instructions at the same time that the hand helper performed the action mentioned in them. In the other condition ("no preview condition"), participants heard nothing through the headphones.

Afterwards, participants were asked to rate their experiences in the experiment, based on different questions, on a 7-point scale from 1 ("not at all") to 7 ("very much"). They were asked the following questions: (i) "To what degree did you feel you could anticipate the movements of the arms?" (ii) "How much control did you feel that you had over the arms' movements?" (iii) "To what degree did you feel you were consciously willing the arms to move?" (iv) "To what degree did the arms look like they belonged to you?" (v) "To what degree did the arms feel like they belonged to you?" And (vi) "Did the arms bother or annoy you?" (Wegner, Sparrow, & Winerman, 2004, p. 840 – 841).

The second and third questions were thought to measure the participant's agentive awareness with respect to the actions of the experimenter's hands. The responses to these questions were correlated, and so Wegner, Sparrow, & Winerman (2004) took the mean of these responses as "an index of vicarious agency" (p. 841). The results showed that the "mean vicarious control ratings" were significantly greater with previews ($M = 3.00$, $SD = 1.09$) than without ($M = 2.05$, $SD = 1.61$). The authors conclude on this basis that

"[i]n line with our hypothesis, the participants receiving previews expressed an enhanced feeling that they were able to control and will the arms' movements" (p. 841).

Just as for the 'I Spy' study, there are significant problems with the interpretation of the results of the "helping hands" study. First, once again, the average ratings given by the participants were very low, even in the preview condition. The average rating was exactly 3, which is less than halfway up the 7-point scale being used. This suggests that participants did not actually feel that they had control over the arms' movements, nor that they felt that they were consciously willing the arms to move. Some theorists have blatantly ignored or misrepresented the low value of the ratings. For example, Synofzik, Vosgerau, & Newen (2008) write:

> In a ''helping hands'' pantomime task, subjects experienced *high degrees* of agency for movements that were in fact performed by another agent, when only the other agent's hands appeared in the place where subjects' hands would normally appear and when subjects could hear instructions previewing each movement. (p. 226, emphasis mine)

But this type of claim, which is not uncommon, dramatically exaggerates the findings of the study.

In addition, Wegner, Sparrow, & Winerman (2004) did not disclose the participants' ratings for each of the individual questions, instead providing only the "combined mean" of ratings on questions (ii) and (iii). And without these, one cannot determine whether, with respect to answers on the individual questions, the ratings were in fact significantly higher in the preview condition than in the non-preview condition. It may be that they were not, but that this finding is buried in the combined mean.

Moreover, the participants may have understood what the experimenters were after, and thereby given increased ratings in the preview condition in accordance with what they believed the experimenters wanted to hear. It is not clear, therefore, that their reports are reliable. Indeed, the authors themselves raise this possibility and offer a reply:

> These findings are suggestive, but they depend on self-report and might conceivably be explained as resulting from participants' guessing the hypothesis and attempting to please the experimenter. Postexperimental questioning (i.e., "Do you have any guesses about what was being studied in this experiment?") indicated that no participant was able to articulate the hypothesis, but further evidence is desirable on the potential role of experimental demand. (p. 841)

But though the participants may not have been able to "articulate" the precise hypothesis of the experiment, they may have had enough of a sense of what the experimenters were after. That is, they might have guessed that the experiment sought to determine whether, when one hears a preview of an action, one rates oneself as having more control over it. And this minimal supposition would likely be enough to contaminate their responses.

Finally, once again, this experiment does not lend any support to Wegner-style high-level matching accounts of agentive awareness. For in this case, too, no intention is involved in the task on the part of the participant. Hearing instructions to A is, of course, not sufficient for forming an intention to A. Similarly, having a thought about doing A is plainly not sufficient for forming an intention to do A—if it were, then there would be no need for practical deliberation. So the 'helping hands' experiment cannot support a matching account of agentive awareness, which requires that one's intention match one's perception of what one is doing.

§3.3 Prime-Effect Studies

Another set of studies that may be thought to support a high-level matching model of agentive awareness uses a paradigm that I will call the *prime-effect paradigm*. In this paradigm, participants perform actions that, on some trials, are followed by effects that are primed beforehand. They are then asked to report the degree to which they felt they had control over those effects.

In one such study, Aarts, Custers, & Wegner (2005) asked participants to move a gray square along a rectangular path of eight white tiles on a computer screen by pressing a "start" key. At the same time, the computer was programmed to move another gray square in the opposite direction at the same speed. When a "stop" signal appeared, the participants were instructed to press a "stop" key to stop their square. The gray squares would disappear in the interval between the signal and the pressing of the stop key. At that time, a *black* square would appear on one of the white tiles, and participants were told that it indicated either the position of their gray square or the computer's gray square.

The twist is that, on some trials, the location of the black square was primed immediately before participants pressed the stop key. The primes were either presented subliminally as brief flashes before the stop signal occurred, or they were presented so that they were consciously seen by the participants. In the latter case, the participants were told to try to stop on the squares that were primed ("conscious goal" condition). The experimenters included this condition in order to determine whether an intention to stop on the relevant square would have an effect on experienced authorship, or whether merely being primed with the location of the square, without forming an intention to stop there, would be sufficient. The participants were then asked whether they caused the

square to land on the spot or whether it was the computer. They indicated their responses on a scale from 1 ("not me) to 10 ("absolutely me"). Aarts, Custers, & Wegner (2005) construed this question as providing "a measure of feeling of authorship" (p. 443).

According to the results, the participants' authorship ratings were significantly higher on trials in which the effect was primed (M = 4.44) compared with those on which it was not primed (M = 3.51). In addition, the higher value of the ratings was independent of whether the prime was subliminal or consciously seen by the participants. Aarts, Custers, & Wegner (2005) conclude that their results "show that feelings of authorship can emerge from subliminally primed prior thoughts of an action's observed effect" (p. 449) and that:

> [s]ubliminal and supraliminal priming were capable of heightening experienced authorship to the same degree as conscious goals to produce these effects. Thus, the observation of behavioral effects can provide the feeling that we caused them when these effects are presaged in our minds. (p. 454)

Again, as in the case of the other studies I have reviewed in this section, these conclusions are too strong given the available data. As in those other studies, the authors conclude that a feeling of authorship was *increased* in the experiment in question by manipulating a key variable, in this case the presence of a prime that previewed the action effect. But a feeling of authorship cannot be increased unless it is present in the first place, and there are a number of reasons to be doubtful that it was in these experiments.

The first is that, once again, the average authorship ratings were towards the mid-point of the scale. If the participants were aware of themselves as stopping their square on the location that the black square appeared, then one would expect them to rate their

authorship towards the high end of the scale. Instead, their mid-range responses seem to indicate that they were unsure whether it was their actions or the computer's programming that had led to the end location of the black square in this ambiguous situation.

Second, the participants did not have control over when the instruction to press the stop key occurred. So on most trials, unless the black square happened to appear on or very near the location the gray square was on when the stop signal occurred, they would be distinctly aware that they did not cause the black square to appear where it did. Moreover, given that they were aware that they had no control over the timing of the stop signal, it is doubtful that they formed any such intention, given that we do not form intentions to perform actions that we do not believe are possible.

Finally, the authors themselves maintain on the basis of their results that an intention is not necessary for agentive awareness to arise—rather, they hold that merely being primed with the effect of one's action is sufficient to produce an experience of authorship over that effect. It is indeed unlikely that the participants formed intentions to land the black square at certain locations, given that they were aware that they had no control over the timing of the stop signal, and this is what determines the location of the black square. Since it is impossible for them to control the outcome of the location of the black square by way of their intentions, it is unlikely that they formed corresponding intentions to do so. But if so, then this study seems to have departed from a focus on agentive awareness, arriving instead at a focus on causal reasoning more generally. One may be aware of oneself as causing an effect without being aware of oneself as the *agent*

of that effect.[38] Suppose, for example, that you are pushed on a crowded subway train, and knocked into someone, causing them to fall. You are aware that you are the cause of the person's fall, but you do not judge that causing him to fall was an action of yours. You would be aware of it as an action, however, if instead you intended to knock him over and pushed him of your own accord. Given that the authors do not take intention or awareness thereof to be playing any role here, it is unclear why this study should be construed as being about agentive awareness at all, which requires that it play at least some role.

In this section, I have reviewed the main experimental evidence offered in favor of a Wegner-style, high-level matching model of agentive awareness. I conclude that the strength of this evidence has been greatly exaggerated. The methodology used in experiments of this type must be suitably refined in order to establish support for the matching model.

§4 The Case Against Wegner-Style Matching Accounts of Agentive Awareness

In addition to the shortcomings I have highlighted in the experimental evidence adduced in favor of Wegner-style matching models, there are also positive reasons *against* adopting such models of agentive awareness. In this section, I will present a case based on theoretical considerations, and one based on empirical considerations. I start with the theoretical case, by canvassing some objections to Wegner's view that do not succeed, before presenting my own that I think is more successful.

§4.1 The Theoretical Case Against Wegner-Style Accounts

[38] Similar studies by Sato (2009) and Wenke, Fleming, & Haggard (2010) suffer from an analogous flaw.

Nahmias (2005) presents an alleged counterexample to Wegner's account from the case of automatic action (see also Pacherie, 2008, p. 203). Nahmias writes that:

> We often experience ourselves as the authors of our well-rehearsed automatic behaviors, as in athletic or musical performances, even when we experience no immediately preceding conscious thoughts (e.g., intentions) to perform the specific behaviors involved. Rather, we seem to have a general intention or plan to play well, perhaps to carry out some array of actions, and then we let our bodies take over, consciously monitoring our movements but not consciously forming intentions to carry out specific subsequent actions (indeed, forming such conscious intentions tends to trip us up). Nonetheless, we experience ourselves as the source of the resulting actions. (p. 774)

At first, one might be tempted to go along with Nahmias (2005) here. Indeed, when we perform skilled or well-rehearsed actions, we typically do not form conscious intentions immediately prior to each stage of the movement. But, on the assumption that Wegner's view requires that we do form such intentions, then we should not expect to have agentive awareness over these actions. It seems that we do, however, so Wegner's account appears to face a difficulty here.

Wegner (2005) is very much alive to this worry, and he offers a potential remedy. He claims that we do *not* in fact typically have agentive awareness over our automatic actions. Instead, he maintains that "many of our most fluid, expert, and admirable acts are ones we do not experience consciously willing,'' resulting in a ''loss of the sense of authorship in skilled actions'' which ''do not feel willed as they unfold'' (p. 29).

Nahmias is not moved by this reply, rejecting it on the grounds that it does not seem to accurately capture our subjective experiences when performing such actions:

> I do not think this is an accurate description of the relevant phenomenology. We do not experience our skilled actions as happening to us or as disconnected from our own intentions, plans, and goals, including some formed during conscious learning and conscious deliberation. Rather, we experience ourselves as the authors and sources of these actions, sometimes the more so precisely because of the earlier conscious effort we have put into them. (p. 783)

While I am sympathetic to Wegner's reply to Nahmias—more on that in a moment—there is a stronger one available to him that blocks Nahmias's objection at an earlier stage. The objection relies on the premise that a Wegner-style model requires that our intentions be conscious. But while it is indeed the case that Wegner himself often puts it this way, there is no reason to think that this is a *necessary* feature of the model. In the case of automatic action, the relevant inference to the effect that one's intention causes one's action could just as easily take place on the basis of nonconscious intentions that guide automatic action—intentions that even Nahmias allows are operative in these cases. There is no reason to insist, then, as Wegner seems to in other places, and as Nahmias does, perhaps echoing Wegner, that in order for one to infer that one's intention caused one's action, the relevant intention must be a *conscious* one. As such, the case of automatic action, arguably involving as it does nonconscious intentions, poses no special difficulty for Wegner's view.

But even if Wegner insists that the relevant intention must be conscious, his reply that we do not have agentive awareness over our automatic actions is adequate for

dealing with Nahmias's objection. Nahmias himself rejects this position on the basis that "we do not experience our [automatic] actions as happening to us or as disconnected from our own intentions, plans, and goals…" (p. 783). However, this is not enough to entail that we have agentive awareness over such actions. Instead, we may simply not be aware of them at all at the time that we are performing them, or be aware of them only as bodily movements, but neither as active nor passive movements. It is true that if we come to subsequently reflect on the question of what we are doing, we often claim full authorship of such actions (though recall the case reported in Marcel, 2003). This does not, however, capture the subjectively unmediated awareness of our actions that agentive awareness in particular involves.[39]

Bayne (2006) presents yet another objection to Wegner's view inspired by a classic example from Davidson (1973/2001c) that is intended to illustrate difficulties with deviant causal chains for causal theories of action.[40] Bayne writes:

> Chloe wants to rid herself of the weight and danger of holding another climber on a rope, and she knows that by loosening her hold on the rope she could satisfy this desire. This belief and want leads her to form the intention to loosen her hold on the rope, but before she has the chance to act on this intention the contemplation of it so unnerves her that she loses her grip on the rope and the climber she is

[39] I will expand on this point in Chapter 5.

[40] The original example from Davidson (1973/2001c) reads as follows: "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" (p. 79). Bayne's (2006) variation adds intentions to the scenario, which Davidson was not yet countenancing as primitive mental states.

> holding falls to his death. Chloe might experience the loosening of her hold as caused by her intention to loosen her hold, but she does not experience the loosening of her hold as an action – she does not experience this event as something that she did. (p. 54)

Bayne takes this example to show that inferring that one's bodily movements are caused by one's intention is not sufficient for agentive awareness. As he puts it:

> … [O]ne can experience one's movements as caused by (and realizing) one's intentions without experiencing a sense of agency towards them. It seems that the experience of doing cannot simply be generated by an awareness of a match between one's intentions and one's movements. Something more is needed. (p. 54)

But this is far too quick. For one, the case of Chloe the climber violates the other two conditions that Wegner specifies for agentive awareness: the priority condition and the exclusivity condition. Recall that Wegner characterizes the priority condition as requiring that the relevant intention must occur *immediately* before the action in question if the agent is to have agentive awareness for that action. But in the case that Bayne presents, Chloe has an intention to loosen her hold on the rope, which is followed by contemplation, which is followed by her being unnerved, which is followed by loosening her grip. So this is not a case in which the intention to loosen her grip occurs immediately before the movement. Rather, there are a number of intermediary events that take place between her intention and her movement.

Second, Wegner's exclusivity condition requires that there be no other salient potential causes of the action in question. But in this case, Chloe would likely be aware

that her becoming unnerved is an alternative candidate for being the cause of her loosening her grip, which might prevent her from inferring that her intention to loosen her grip is the cause of her doing so. And without this inference she would not, on a Wegner-style view, have agentive awareness. As such, the case cannot be used to undermine such an account, as it would itself predict that there is no agentive awareness in this case. Bayne appears to be treating the view as holding that a match between intention and action is both necessary and *sufficient* for agentive awareness, but this is not the case.

While the criticisms I have been surveying fail, however, there is one that succeeds. The problem with Wegner's view is that it assigns far too heavy a role for causal reasoning with respect to a process that is, in mature and healthy agents, sufficiently fine-tuned and reliable so as not to require such reasoning to take place. When it comes to the vast majority of our actions—and we perform myriad each day—the causal relationship they bear with our intentions is such that we do not need to establish it anew each time, though we may need to do so under certain circumstances. This is a simple point, and yet somehow it has not been properly appreciated when it comes to evaluating Wegner-style accounts of agentive awareness. I expand below.

Wegner is doubtless correct in viewing the type of causal reasoning he highlights as central to *many* of our action attributions. Consider a version of what, following Feinberg (1970), has come to be known as *the accordion effect*. Suppose an agent S performs an action A. If A causes event B, then we infer that B is an action that S performs—and this applies to both first-person and third-person action attributions. Davidson (1971/2001b) sums the effect up in the following slogan: "An agent causes

what his actions cause" (p. 53).[41] Moreover, the causal inferences involved in the accordion effect are arguably sensitive to the conditions that Wegner lays out of priority, exclusivity, and consistency—except that they are not inferences about the causal relationship between intentions and their effects, but about bodily actions and *their* effects. The accordion effect thereby illustrates how action attributions extend from one's bodily actions to the subsequent events that one's bodily actions cause, and the role of causal reasoning in yielding these attributions.

But, importantly, the accordion effect has its limits. Thus, Davidson (1971/2001b) writes that: "The ordinary notion of event causality is useful in explaining how agency can spread from primitive actions to action described in further ways, but it cannot in the same way explain the basic sense of agency" (p. 52). In other words, Davidson denies that the same type of causal reasoning underlies action attributions pertaining to basic or primitive actions, that is, roughly, actions that do not require that one does something else first (see Danto, 1965). When it comes to actions like raising one's arm, or flexing one's knee—assuming these are basic actions—one need not resort to causal reasoning about the relationship between one's intention and one's bodily movement. Action attributions in such cases, Davidson maintains, operate by way of a different mechanism—though he does not specify what it is.

Davidson's (1971/2001b) main reason for denying the contribution of causal reasoning in such cases is that he rejects intentions as primitive states, which could

[41] As Searle (1983) nicely put it with respect to Gavrilo Princip, the archduke Franz Ferdinand's assassin: "It is a remarkable and little-noted fact of human and animal evolution that we have the capacity to make intentional bodily movements where the conditions of satisfaction of our intentions go beyond the bodily movements. Princip moved only his finger but his Intentionality covered the Austro-Hungarian Empire" (99).

thereby anchor one end of a causal relationship. He views them, rather, as pairs of beliefs and pro-attitudes towards an action, by appeal to which actions are classified as intentional. But there is good reason to agree with Davidson with respect to the limits of the accordion effect, independent of his eliminativist approach to intentions. In what follows, I will make the case for an alternative proposal as to why the causal reasoning fails to capture the mechanism underlying action attributions, not just pertaining to *basic* actions, but to bodily actions more generally.

As I remarked earlier, there is an important asymmetry with respect to the causal relationship that holds between our intentions and our bodily movements, on the one hand, and our bodily movements and subsequent non-bodily events, on the other. Unlike the causal relationship between many of our bodily movements and non-bodily events, the causal relationship between our intentions and our bodily movements is, for healthy and mature agents acting in normal circumstances, robust, reliable, and well-learned. When one forms an intention to perform an action, except in unusual circumstances—e.g., temporary paralysis, deficits in motor control, novel or unfamiliar contexts—the appropriate bodily movement takes place straightaway. What this means is that, typically, the question of whether one's intention is the cause of one's bodily movement *simply does not arise*—and so there is typically no role for causal reasoning to play here.

Wegner's (2002) model is at its most plausible when applied to cases in which the causal relationship between one's intention and one's action is *not* reliable or not *judged* to be reliable. In such cases, it may indeed be necessary, in order to become aware of oneself as performing the relevant action, to rely on the markers of causality that Wegner specifies—that is, to "check up" on the efficacy of one's intentions. Wegner's empirical

work, for example the 'I Spy' experiment, emphasizes this objective, third-person stance towards one's own actions, involving as it does the connection between one's jointly moving a mouse and a cursor on a computer screen—a connection of weak reliability given the contributions of the other agent.

But, once again, this is misleading in an important way. For it mistakes the *exception* for the *rule* when it comes to one's *first-person* agentive awareness. Typically, because of the reliable connection just highlighted, one does not need to engage in causal reasoning of the sort that Wegner describes, conscious or not, in order to judge that one is performing some action, and for agentive awareness to arise. In the next chapter, I will elaborate on the details of how agentive awareness arises in the absence of Wegner-style causal reasoning. For now, suffice it to say that Wegner is operating on a misguided model of agentive awareness that inappropriately lumps it in with causal judgments more generally.

§4.2 The Empirical Case Against Wegner-Style Accounts

In this section, I present two cases that together provide compelling evidence against Wegner's account of agentive awareness, and high-level, matching models more generally. More specifically, they suggest that a match between an intention and one's action is not necessary for agentive awareness to arise.

§4.2.1 Deafferented individuals

Deafferentation is a condition in which an individual displays significant impairments in touch and prioprioception. There are cases of deafferented individuals who have lost almost all sense of proprioception, the most famous of which are Ian Waterman (see Cole, 1995) and GL (see Fourneret, Paillard, Lamarre, Cole, & Jeannerod,

2002). At the age of 19, Waterman was struck with what appeared to be gastric flu, believed to be caused by an autoimmune reaction, which led to deafferentation from the neck down. Waterman, at first, lost all control of his body, being unable to sit or stand up, walk, feed himself, or interact with ordinary objects. But after years of rigorous practice and rehabilitation, he is now able to guide his body with the help of vision, and he is remarkably good at it (Cole, 1995).

GL is a 55-year-old woman from Québec who, at the age of 31, suffered carbon monoxide poisoning, resulting in sensory neuropathy. She was left with deafferenation from her nose to her toes (Guillaud, Simoneau, & Blouin, 2011). In her case, too, intensive training and rehabilitation has helped her to regain visually-guided control over her body.

Both Ian Waterman and GL have participated in numerous studies aimed at elucidating the nature of their condition and the extent of their deficits. Most of this research focuses on the implications for our understanding of sensorimotor control (e.g., Guillaud, Simonau, Blouin, 2011; Sarlegna & Bernier, 2010; Sarlegna, Gauthier, Bourdin, Vercher, & Blouin, 2006). But there is little, if any, explicit attention paid to the question of whether and how these individuals undergo agentive awareness.

Still, some insights can be gleaned from their performance on sensorimotor control tasks and from their verbal reports. Care must be taken, however, in how these reports are interpreted. For example, in discussing the case of another deafferented individual, Christina, as described by Sacks (1998), Nahmias (2005) arrives at the unwarranted conclusion that Christina reports not having agentive awareness over her actions. Nahmias takes Christina's case to be a counterexample to a Wegner-style

matching model, not with respect to the necessity of a match between one's intention and sensory feedback, as I am presently arguing, but with respect to its sufficiency. More specifically, he takes issue with the claim that all three of Wegner's conditions are sufficient for agentive awareness to arise. Concerning the case of Christina, Nahmias writes:

> Eventually she learns to get around by coordinating her intentions with her visually observed movements—her intentions are then consistent with, temporally prior to, and perceived to be the exclusive causes of her actions. Yet, despite meeting these conditions of Wegner's model, she does not feel her normal sense of agency or self nor move as fluidly or efficiently. As she puts it, ''It's like something has been scooped right our of me, right at the centre . . . see Chris, the first pithed human being [with] no sense of herself—disembodied Chris, the pithed girl!'' […] Christina's loss of the experience of her own body seems to disrupt her experience of herself as the source of her actions, even though she has the relevant experiences of priority, consistency, and exclusivity sufficient to infer that her intentions are the cause of her bodily movements. (p. 773)

Nahmias is not careful, however, to distinguish between awareness of oneself as embodied, and awareness of oneself as an agent. But Christina's remarks are only about the former, and not the latter. She feels she lacks a sense of herself much of the time arguably because she does not feel her body, so her bodily self is impaired. This is not the same as lacking agentive awareness, however.

Indeed, there is some evidence that points towards deafferented individuals having agentive awareness over at least some of their bodily actions. For example,

commenting on GL's performance in an experimental task, de Vignemont and Fourneret (2004) note that: "She did not pretend to have moved while she did not, like anosognosic patients, nor did she make the same movement twice believing that she had not moved yet" (p. 13). This would suggest, on the face of it, that GL has awareness of herself as acting. And if GL, Ian Waterman, and other deafferented individuals sometimes have agentive awareness over what they do, then it would seem that this is a counterexample to Wegner's account, and to high-level matching models more generally, which require sensory feedback from one's movement in order for agentive awareness to arise.

Carruthers (2012) takes issue with this line of reasoning, arguing that deafferented individuals do not pose a problem for matching models of agentive awareness. He urges that, though proprioceptive feedback may not be required in order for such individuals to have agentive awareness, at least some of the time, they do require *visual* feedback. He writes:

> … it seems these patients do not experience a sense of agency over actions they do not see, at least for the case of conversational gestures. Both of these patients [IW and GL] gesture when talking. G.L. was shown a video of her doing so. On seeing this she reported that she did not feel that she controlled the movements at the time they occurred… (p.33)

Carruthers is arguing here that without vision, GL reports not being in control of her movements, which would seem to indicate that she does not, in fact, possess agentive awareness over actions she does not see.

While this may seem to dismiss deafferented individuals as a potential counterexample to high-level matching accounts, Carruthers (2012) does acknowledge

that the case he presents is not conclusive, and I agree. For one thing, GL's report of not feeling in control of her movements in the video is compatible with her not being aware of making them at all, something that commonly happens when we move in automatic or habitual ways. Although the conversational gestures of Waterman and GL began as forced, effortful, and deliberate in the early stages of their shared condition, after much training and effort they have each succeeded in making them automatic and effortless (Cole, Gallagher, & McNeill, 2002).

Carruthers (2012) discusses another case involving Waterman, which he takes to suggest that deafferented individuals do not have agentive awareness over conversational gestures that they do not see. Waterman was asked to narrate the plot of a cartoon while wearing a blindfold. Twenty seconds in, after he had already produced 14 conversational gestures, Waterman reportedly said, "… and I'm starting to use my hands now…" suggesting that he had not been previously aware of doing so, which is what would be predicted by a matching model like Wegner's that requires perception of what one is doing.

Still, Waterman's report that he is starting to use his hands now suggests that he *does* have agentive awareness at *that* time despite the absence of proprioceptive and visual feedback. Carruthers (2012) denies this:

> We may ask if his claim that he is beginning to act (I'm starting to use my hands) arises from a sense of agency. Did his sense of agency 'kick in' 20 s after he began his action? It is unlikely. I.W. could not be reporting a feeling that his hands are starting to move, as he has no way to experience that. More likely this report comes from knowledge of his intention to begin moving. In other words

> this report is a result of I.W.'s theory of mind and not his sense of agency. As such it seems most plausible that I.W. does not experience a sense of agency for these actions. (p. 34)

Carruthers claims here that if Waterman's awareness of performing an action stems from awareness of his intentions, then it is not agentive awareness, but something else. This is, however, to beg the question against an account on which agentive awareness can be had in the absence of sensory feedback. Such an account might very well appeal to intentions as an alternative source of agentive awareness, as I will go on to sketch in Chapter 5.

It remains mysterious, however, why Waterman would come to be agentively aware 20-seconds after having started to move, given that he would have formed a volition to do so much earlier. In general, the literature on deafferented individuals suggests more questions than answers when it comes to agentive awareness. I raise this type of case here to flag its relevance to this kind of research, while recognizing the limitations of its evidentiary force given the current state of the literature.

§4.2.2 Libet's M-judgments

In a classic study, Libet, Gleason, Wright, Pearl et al. (1983) set out to determine the relationship between one's willing and initiating an action, and one's awareness thereof. Libet et al. asked six participants each to perform a series of simple actions—forty in total. More specifically, they were instructed in each trial to perform a "quick, abrupt flexion of the fingers/and or wrist" (p. 625) with their right hands, at a time of their own choosing. They were encouraged not to decide in advance on a time at which to perform the act, but rather "to let the urge to act appear on its own at any time without any preplanning or concentration on when to act" (p. 625). Participants were seated

facing a specialized clock, around which a dot would revolve every 2.56 seconds. They were instructed to report the time at which they first became aware of deciding or having the "urge" to move based on the position of the dot on the clock ("W judgment"). They were also asked to report the time at which they were aware that they had "actually moved" (p. 627) ("M judgment"). At the same time, they were hooked up to an electroencephalogram (EEG), which measured their brain activity during this these tasks.

Much of the discussion surrounding this study has focused on the striking finding that a neural event, the main negative shift of the Readiness Potential (or RP), which is associated with the initiation of action (see also Haggard & Eimer, 1999 for a refinement of the results; though see Schurger, Sitt, & Dehaene, 2012), precedes the time at which participants report being aware of an urge to act by about 350 ms. The reported time of the urge itself precedes the onset of the movement by about 200 ms. Some have worried that this poses a threat for free will, since it suggests that actions are initiated nonconsciously in the brain, and a debate has raged for some time now regarding the validity of this concern (see Rosenthal, 2002 for discussion).

While the debate surrounding the timing of the W-judgment and its relation to free will is intriguing, for present purposes, the findings related to participants' M-judgements are of greater interest. The results here were that participants reported being aware of acting an average of 86 ms *prior* to the onset of muscle activation as measured by an electromyogram (EMG). So before there was any movement at all, according to the participants' reports, they were aware of themselves as acting. Moreover, this result has been widely replicated; anticipatory awareness of action has since been found in a

number of studies using similar paradigms (Haggard & Eimer, 1999; Haggard, Newman, & Magno, 1999; Lau et al., 2004; Rosenthal, 2002).

The upshot here is that people seem to be aware of themselves as acting prior to sensing any bodily movement. The individuals in these studies are aware of performing actions, that is, agentively aware, before the arrival of sensory feedback from their actual movement—indeed prior to any movement whatsoever. This strongly suggests that agentive awareness does *not* require a match between one's intention to A and one's perception of A-ing. The awareness, at least sometimes, arises independently of any such a process, which goes against the predictions of Wegner's account, and any matching model more generally. The results of Libet's classic study thus furnish a compelling counterexample to such models of agentive awareness.

There are those who take issue with Libet's study on the basis of methodological concerns with his experimental paradigm or his interpretation of his results (e.g., Schurger et al., 2012). But these criticisms mainly pertain to the W-judgement trials, and not the M-judgment trials that are my present focus. Some critics, however, also address the latter trials. For example, in an elegant study, Lau, Rogers, & Passingham (2007) suggest that participants' M-judgments are sensitive to arbitrary factors that should make no difference if participants are really basing them on the time at which they become aware of moving. Using Libet's paradigm, Lau and his colleagues asked participants to perform spontaneous button presses, and then to move a cursor to the position on the clock that the revolving dot was in when they pressed the button. In half the trials, TMS was administered over the pre-SMA, and in the other half, a "sham TMS" was activated, which involved another TMS machine with a coil placed near the back of the head, but

away from the cortex. The purpose of the sham TMS trials was to prevent participants from knowing during which trials the pre-SMA area of their brain was being stimulated. The real or sham stimulations were applied either simultaneously with the participant's action or 200 ms afterwards.

Lau, Rogers, & Passingham (2007) found that TMS stimulation over the pre-SMA induced a forward shift in the reported awareness of the onset of action regardless of whether it was administered at the time of action onset, or 200 ms later. When the TMS was applied at the time of the action, there was a 14 ms delay in the time reported in the M-judgment relative to that reported in the sham TMS trials. In other words, participants reported being aware of acting 14 ms later than they reported acting when only a sham TMS was administered. And when TMS was applied 200 ms after movement, there was still a 9 ms delay in the reported timing of the action relative to that reported in the sham TMS trials. These effects, though small, were found to be significant—the application of TMS had the result of distorting participants' M-judgments (for similar TMS manipulations to M-judgment, see also Haggard, Clark, and Kalogeras, 2002; Haggard and Magno, 1999).

It may be tempting to interpret these results as casting doubt on the reliability of participants' M-judgments. Perhaps, one might argue, the results suggest that these judgments do not accurately reflect the time at which participants first became aware of themselves as acting, but are rather based on cues occurring *after* the movement. After all, TMS administered 200 ms after the movement still had an effect on these judgments, and if so, then perhaps they do not reflect agentive awareness in the absence of a match with a perceptual state, as I have claimed.

This interpretation of the results is too extreme, however. It does not follow from the fact that participants' M-judgments can be interfered with after the action that, in the absence of such interference, they are not reliable. This would be analogous to arguing that backwards masking shows that the relevant stimuli are not perceived consciously when a mask is *not* present. Or, it would be like arguing that the fact that one's attention to a task can be distracted by a loud noise is evidence that one does not properly attend in the absence of any loud noises. Each of these inferences is problematic, as they take performance in suboptimal conditions as a guide to performance in regular circumstances.

Some further resistance to the interpretation of the M-judgment as offering up a counterexample to matching models comes from de Vignemont and Fourneret (2004). They argue that agentive awareness is divided into two components: (i) the sense of initiation, and (ii) the sense of one's own movements. The sense of initiation involves awareness of starting to act, whereas the sense of one's own movements involves awareness of the *way in which* one is acting, for example, what specific limb movements are employed. De Vignemont and Fourneret (2004) urge that what is reported on in participants' M-judgments in Libet-style paradigms is the sense of initiation, not the sense of one's own movements, and so these judgments cannot be taken as indications of full-fledged agentive awareness, involving as it does *two* components.

It is doubtful, however, that agentive awareness can be coherently bifurcated in this manner. Awareness of starting to move is also thereby awareness of moving *in a certain way*—that is, awareness of the way in which the action is achieved. There is no bald awareness of moving that does not in some way, however loose, specify how it is

that one is moving. It is likely that participants in Libet's paradigm are not merely aware of initiating an indeterminate movement, but of initiating some *particular* movement, e.g., a flexing of the wrist.

But de Vignemont and Fourneret (2004) argue that the sense of initiation and the sense of what one is doing do indeed come apart in certain cases, thereby indicating that they are distinct components of agentive awareness. They claim that anosognosic patients, for example, have an impaired sense of initiation but not an impaired sense of what they are doing. Describing a particular case study, they write:

> When the patient suffering from anosognosia squeezed with his normal hand, both the right and the left pectoralis contracted. But when asked to squeeze with the contralesional hand, he did not contract either pectoralis muscle. Therefore, even if the movement—or at least the required muscle contractions—could still be triggered automatically, the patient was no longer able to voluntarily initiate a movement. However, he was not conscious of this deficit and felt as if he were really initiating the movement when he wanted. We can thus conclude that anosognosia is associated to the disruption of the sense of initiation. (p. 12)

What is described here is not a case of an impaired *sense* of initiation, though, but rather of impaired initiation itself. After all, de Vignemont and Fourneret (2004) allow that the individual "felt as if he were really initiating the movement," despite not moving. This constitutes awareness of initiation, which is precisely what is at issue here.

The second case that de Vignemont and Fourneret (2004) present fares no better at establishing a dissociation between the sense of initiation and the sense of one's own movements. They claim that the deafferented individual GL has an impaired sense of

what she is doing, but not an impaired sense of initiation. Describing GL's performance in a particular study, they write:

> In a recent study with this patient using the 'alien hand' task, Farrer et al. (personal communication) also revealed that she self-attributed visually presented movements even if the visual feedback was strongly biased. However, she perfectly knew that she was moving. […] Therefore, the sense of initiation is preserved whereas the sense of one's own movements is impaired. (p. 13)

Again, it is not clear that we can accurately say of GL that she is aware of having initiated some movement, but not aware of which one. It may be true that she is unable to track the precise spatial or temporal properties of her action as it unfolds, given the lack of proprioceptive feedback, but this is not the same as saying that she has *no* awareness of what she is doing, which is what is required to establish that a sense of what one is doing can come apart from a sense of initiation.

In sum, there is no evidence that one can be aware of oneself as initiating a movement, but not be aware in any way of which movement one has initiated. And if so, then there is no reason to hold that agentive awareness is best understood as involving two separable components, only one of which is the sense of initiation. This means that we are free to view participants' reports in the Libet M-judgment task not just as reports on the awareness of *initiating* action, but on acting more generally—as reports of genuine agentive awareness.

Carruthers (2012) also takes issue with viewing the M-judgment results of Libet-style studies as providing a counterexample to matching models of agentive awareness. He writes:

> There is a problem with interpreting the reports of the subjects in this study. In this study subjects were asked to recall when they started to move. However, there is no way to tell at this later time if subjects have an experience of moving initially involving the sense of agency or the experience not initially involving a sense of agency, but having the sense added later… The second hypothesis would merely involve the sense of agency being reported as occurring earlier than it actually did. As we have no way to differentiate between these two versions of events based on the subjects reports alone, the above argument cannot establish the claim that the actual sensory consequences of an action cannot be involved in eliciting the sense of agency (p. 32)

On Carruthers' alternative interpretation, when the participants are first aware of moving, they are nonetheless not aware of themselves as performing an action, that is, they lack agentive awareness. This awareness comes later and is subsequently reported as occurring earlier than it did.

But why think that when participants are first aware of moving, this awareness is not agentive? If participants are not aware of their movement as being their action, then either they are aware of it as a passive movement, or they are aware of it in a way that is neutral between its being a passive movement or an action. It is unlikely that they are aware of it as being a passive movement, since they are told to perform a simple action, and these are typically, in healthily functioning individuals, accompanied by agentive awareness. There is no reason to prefer an interpretation on which they are not agentively aware of their movements, apart from to save this objection. For the same reason, it is also unlikely that participants are aware of their movement, but as being

neither agentive nor passive. This would be a highly unexpected result given the circumstances, and we do not have a compelling reason to accept it.[42]

§5 Conclusion

An overwhelming majority of theorists working on agentive awareness endorse a matching model of such awareness. In this chapter, I have offered a tripartite challenge to high-level matching models of the type defended by Wegner (2002). First, the empirical evidence that is often appealed to in support of such models does not hold up. Second, the matching model assimilates all agentive awareness to a form of causal reasoning. But though causal reasoning may be operative in some agency-attributions it is not the underlying mechanism of agentive awareness. Third, there are cases in the empirical literature—deafferented individuals and anticipatory agentive awareness in Libet-style studies—that seem to suggest that one can have agentive awareness in the absence of any match between an intention and an action.

[42] It is worth stressing that the two counterexamples I have presented here against the matching model also serve as counterexamples to sensory models of agentive awareness. If I am right, these are cases in which an individual is agentively aware in the absence of any sensory feedback whatsoever—so agentive awareness cannot be a form of sensory awareness. They also serve as counterexamples to the CM approach.

CHAPTER 5: A COGNITIVE ACCOUNT OF AGENTIVE AWARENESS

§1 Introduction

I would now like to take the lessons learned in the last three chapters and use them as guides in developing my own theory in this chapter. In Chapter 2, we saw that agentive awareness is not a form of sensory awareness—it is had in virtue of non-sensory states. In Chapter 3, I argued that whatever types of states these are, they are not low-level states of the kind involved in sensorimotor control, such as those posited by the comparator model. In Chapter 4, I argued against matching models of agentive awareness. We are left then to develop a non-sensory, high-level, non-matching account of agentive awareness. I turn to this task in this chapter.

Given these guidelines, a natural place to look is towards an account of agentive awareness on which intentions play a major role. Following other theorists, we may distinguish between two kinds of intention here (Brand, 1984; Bratman, 1987; Mele, 1992; Pacherie, 2008; Searle, 1983). In some cases, one forms an intention to do something in the future. One may intend to take a walk later, or attend a concert tomorrow, or go on vacation next year. Following Mele (1992) and Pacherie (2008), I will call these *distal intentions*. In addition to intending actions for the future, however, one may also intend to do something *now.* One may intend to go for a walk now, or make a cup of coffee now, or call a friend now. I will call such intentions *executive intentions,* in order to highlight their role in the initiation and guidance of present action, rather than

the planning of future action.[43] Given this primary role, I understand executive intentions as belonging to the same family as what other theorists call *volitions* (Sellars, 1976), *proximal intentions* (Mele, 1992; Pacherie, 2008), and *intentions-in-action* (Searle, 1983). Though there are differences among such states when one gets down to the details, they will not concern us here.

Distal intentions are plainly not suitable states for making one agentively aware. They are formed too far in advance of action to give an individual information concerning what she is presently doing. For this we need to turn to executive intentions. Later on it will emerge that executive intentions are not by themselves likely to be sufficient for agentive awareness, though they will play a key role in generating it. Nonetheless, they provide a useful starting point for theorizing.

An intention-based account of agentive awareness holds much plausibility. After all, actions are events that are caused in the right way by intentions. And if so, then intentions are a rich source of information about what one is doing. It would then be no surprise if, as agents, we regularly draw on this source. Indeed, some of the experimental work I have already discussed supports this claim. Recall the Lau, Rogers, & Passingham (2007) study from Chapter 4, which found that TMS stimulation over the (pre-)SMA results in a delay in the reported awareness of acting, both in cases where the stimulation was applied at the time of action onset, and 200 ms later. I argued against construing this study as reason to view participants' M-judgments as unreliable, since many judgments can go astray in the face of interference, despite their being reliable under ordinary

[43] Ginet (1990) nicely captures the character of such intentions in his remark that "they do not plan ahead, not even very slightly. They do not *plan* at all; they *execute*" (33).

circumstances. But, furthermore, Lau et al.'s own interpretation of their results actually helps to support the view that agentive awareness is somehow based on executive intentions.

Lau, Rogers, & Passingham (2007) offer a Bayesian cue integration model to explain why TMS over the (pre-)SMA results in a delay in agentive awareness. On that model, in order to arrive at a judgment as to when one is acting, one uses information from various action-related cues. There are two main cues available in the case with which they are concerned: (i) the executive intention to press the button now, and (ii) the proprioceptive feedback from the movement. There is growing evidence that executive intentions are neurally realized in the SMA. Stimulation in this area (medial frontal cortex) elicits reports of experiencing a conscious urge to move (Chassagnon, Minotti, Kremer, Hoffmann, & Kahane, 2008; Fried et al., 1991; Lafargue & Franck, 2009; Pacherie & Haggard, 2010). In addition, lesions to this area (medial frontal cortex) result in the complete impairment of self-initiated movements in macaque monkeys (Thaler, Chen, Nixon, Stern, & Passingham, 1995). And when people are asked, in a Libet-style paradigm, to attend to the timing of their decision to move, activity in the pre-SMA increases (Lau, Rogers, & Passingham, 2006).

Stimulating a brain region with TMS adds noise to signals produced in this region, so that they are less reliable (Lau, Rogers, & Passingham, 2007). This means that the awareness of acting will need to depend instead on the proprioceptive feedback from the movement. As such the awareness of having acted is judged to occur later in time than in the non-TMS trials. But what this *further* suggests is that in *typical* cases, where

no TMS interference occurs, the awareness of having started acting in some way depends on executive intention itself, located as it is in the area that is disrupted by the TMS.

Still, there are a number of ways an intention-based account of agentive awareness might work. In this chapter, I will present two main alternatives, before settling on one in which one is agentively aware in virtue of having a thought to the effect that one is A-ing that is based on an executive intention to A. I start, however, by providing a defense for positing executive intentions in the first place, as well as a sketch of their functional profile.

§2 In Defense of Executive Intentions

§2.1 Some initial challenges

Executive intentions, though more commonly posited in action theory these days, have not always been embraced in the philosophical literature. Indeed, in *The Concept of Mind*, Ryle (2000) famously lampoons the concept of volition (or executive intention) as hopelessly obscure and lacking in utility—one that belongs in the historical dustbin along with the concepts of 'phlogiston' and 'animal spirits'. It will be useful to start out my defense of executive intentions with a takedown of some of Ryle's classic challenges.[44]

One worry Ryle raises about executive intentions is that they do not figure in folk psychology. According to Ryle, these "ghostly thrusts" (65) play no role in everyday descriptions of either one's own actions or those of other agents, in our legal discourse,

[44] Of course, in presenting these challenges, Ryle has a particular agenda in mind, namely to dismantle what he calls the "Official Doctrine", which champions the central tenets of dualism. On a standard interpretation of Ryle's project, he aimed to give an analysis of mental vocabulary terms such as 'volition', 'pain', and 'desire' wholly in terms of behavior and dispositions to behave. Evaluating this project is not here my focus, so I will be distilling his arguments against volitions from his more general purposes in the present treatment.

nor in literary descriptions—all the places that one would expect a folk psychological concept to appear. As Ryle puts it:

> No one ever says such things as that at 10 a.m. he was occupied in willing this or that, or that he performed five quick and easy volitions and two slow and difficult volitions between midday and lunch-time. An accused person may admit or deny that he did something, or that he did it on purpose, but he never admits or denies having willed. Nor do the judge and jury require to be satisfied by evidence, which in the nature of the case could never be adduced, that a volition preceded the pulling of the trigger. Novelists describe the actions, remarks, gestures and grimaces, the daydreams, deliberations, qualms and embarrassments of their characters; but they never mention their volitions. They would not know what to say about them. (p. 64)

The upshot of this challenge is, presumably, that if executive intention does not play any useful explanatory role in our folk psychology, then this is all the more reason to eliminate it in favor of theoretical entities that *do* have explanatory power.

Ryle does not succeed, however, in establishing that executive intention is not a folk psychological concept. First of all, some of his complaints apply just as well to beliefs and desires, which few would deny are included in folk psychology, as they do to executive intentions. One never says that "at 10 a.m. he was occupied in" believing or desiring "this or that", or that "he performed five quick and easy" beliefs or desires "between midday and lunch-time". But this does not undermine the role of beliefs and desires in folk psychology.

Furthermore, the fact that people do not typically appeal to their executive intentions in describing their behavior has a ready alternative explanation. It is a key feature of executive intentions that they are intimately tied to the occurrence of bodily actions. Only in extremely rare circumstances, does one form an executive intention that does not immediately result in a corresponding action. And, as bodily actions are overt events that are publicly observable, it is almost always more appropriate to speak of the actions themselves and their properties, rather than the intentions that cause them—and nothing is typically lost in doing so.

The relatively rare times in which something is lost in describing actions and properties rather than the mental states—including intentions—that cause them are when the action does not go as intended, and so does not correspond to the relevant mental states. But in these cases people *do* appeal to their intentions to explain the character of the action—saying things like "I didn't mean to do that," or "I didn't do that on purpose," as Ryle himself later acknowledges. It is unclear, then, why Ryle denies we talk about executive intentions in the way that would be expected if they were folk constructs.

Ryle presents another challenge to the positing of executive intentions in the form of a dilemma. He maintains that executive intentions are posited to explain "that which makes actions voluntary, resolute, meritorious and wicked" (p. 67), but these predicates apply to mental events, such as imagining, no less than bodily events. So, Ryle asks, given that executive intentions are themselves mental events, are they voluntary or involuntary? He takes each horn of the dilemma to lead to "absurdities" (p. 67), complaining that:

> If I cannot help willing to pull the trigger, it would be absurd to describe my pulling it as 'voluntary'. But if my volition to pull the trigger is voluntary, in the sense assumed by the theory, then it must issue from a prior volition and that from another *ad infinitum* (p. 67).

Ryle rejects the first horn of the dilemma because, he thinks, it is absurd to hold that something involuntary, that is, a "willing" that one "cannot help", could cause something voluntary, like an action. But he offers no reason to support this claim and, indeed, there is no good reason to suppose that causes must share similar properties with their effects—this would be to endorse a view of causality on which causes *transfer* some property to their effects. And there is no reason to adopt such a view.

Ryle rejects the second horn of the dilemma because he takes it that, if an executive intention is voluntary, then in order to explain its occurrence, one must posit another executive intention, which itself must be explained by yet another executive intention, and so on. But Ryle starts off on the wrong foot to begin with. Executive intentions are indeed mental events, but they are not actions, any more than beliefs and desires, being themselves mental events, are actions. There is no need, therefore, for executive intentions to be preceded by yet other executive intentions in order to explain their occurrence. And since there is no problem about something that is not voluntary, or not an action, causing a voluntary movement, or an action, there is no issue about executive intentions causing voluntary movements or actions either.

§2.2 Why posit executive intentions?

While Ryle's attempts to undermine the concept of volition are ultimately unsuccessful, they do, as a whole, get at an important question: What specific reason or

set of reasons do we have to posit executive intentions? In other words, what feature or features of agency do we need executive intentions to *explain*?

Searle (1983) made popular an argument for positing executive intentions that I will call the *argument from indeterminacy*. The argument runs as follows: Most if not all of our intentional actions involve "subsidiary actions" that are not specified by any distal intention, but that are nonetheless performed intentionally. To borrow Searle's example, suppose one forms an intention to drive to one's office. In proceeding to do so, one performs a number of actions, such as shifting from second gear to third gear, and pressing the gas pedal. But one does not typically form any distal intentions—either conscious or not—to perform any of *these* actions—one simply intends to drive to one's office. Nonetheless, they are intentional.

Since we cannot explain the intentional character of these more basic actions by way of any distal intentions, we must do so by way of positing another type of intention that is operative in this case: an executive intention. So in addition to a distal intention to drive to one's office, one forms an executive intention for every shift in gear, every pedal press, and every other action that one performs in the service of this more general intention. Or so the reasoning goes.

This argument rests on a contentious premise. The premise is that in order for one's action of shifting from second to third gear, for example, to be intentional, one must form an intention—of one type or another—to shift from second to third gear. In other words, it rests on the claim that in order for one to intentionally A, one must intend to A. This is the main thesis of the view that Bratman (1984) has dubbed *The Simple*

View, and there has been much debate in action theory surrounding its validity (Bratman, 1984, 1987; Mele, 1992).

For example, many urge that our intuitions in cases in which there are foreseen but unintended side-effects of action yield convincing counterexamples to the main thesis of The Simple View. Harman (1976) offers the following well-known case of a sniper who, in firing his gun, intentionally alerts the enemy despite not intending to do so:

> In firing his gun, [a] sniper [who is trying to kill a soldier] knowingly alerts the enemy to his presence. He does this [i.e., alerts the enemy] intentionally, thinking that the gain is worth the possible cost. But he certainly does not intend to alert the enemy to his presence. (p. 433)

And others have shared Harman's own intuitions in response to this case and others like it.[45]

But intuitions, no matter which way they point, cannot resolve this issue on their own. If they are pre-theoretical folk intuitions, then there is evidence from experimental philosophy that they are unreliable. Indeed, there is mounting evidence that intuitions, as they figure in just these types of side-effect cases, are sensitive to irrelevant features of the scenario, such as whether the action in question is morally blameworthy or praiseworthy (Knobe, 2003). And if they are not pre-theoretical folk intuitions, but instead "expert" intuitions from people well-practiced in philosophical theorizing, then they are themselves likely just deliverances of implicitly held theories, and so they cannot then play a neutral role in settling theoretical matters (see also Rosenthal, 2010, p. 372).

[45] See also Bratman's (1987) case of a marathon runner who intends to run a marathon and expects to wear down her shoes as a side-effect. Here, too, many have the intuition that she wears down her shoes intentionally, despite not intending to do so.

But intuitions aside, there is no reason to hold the Simple View. Rather, we can explain why an action is intentional by appeal to whether it is appropriately guided by an agent's intention, that is, by appeal to whether it is under the agent's intentional control. And, for an action of A-ing to be intentional, it need not be an intention to A that does the guiding—it may instead be a distal intention to do something to which A is related as a means, that does the guiding. The point is that intentional actions must be related to intentions in appropriate ways, not that they must in every case be specified in the content of a preceding intention. So the argument from indeterminacy does not give us good reason to posit executive intentions in addition to guiding distal intentions.

Another argument for positing executive intentions may hit closer to the mark. I will call it the *argument from spontaneous action*. The argument runs as follows: Sometimes one performs an intentional action "on impulse", without any time to form a distal intention to act. For example, one might suddenly strike someone who says something insulting (Searle, 1983), or slam on the brakes in response to an animal running onto the road (Mele & Moser, 1994). Applied to executive intentions, the argument is that we must posit them to explain the *intentional character* of spontaneous actions, since all intentional actions must at least be intended, and there is no distal intention here to do the relevant work.

This argument does fare better than the argument from indeterminacy, since it is committed to the claim that in order to A intentionally, one must intend to do something, though not necessarily to A, and this holds much plausibility. Still, one might resist this line of reasoning, and insist that spontaneous actions *are* preceded by nonconscious distal intentions. Indeed, there is a way of interpreting recent empirical work by Soon, Brass,

Heinze, & Haynes (2008) that seems to point in favor of this hypothesis. Soon et al. (2008) found that brain activity in the prefrontal and parietal cortex occurring up to 10 seconds before a spontaneous action (pressing a button with either the left finger or the right finger) was predictive of that action. One way of construing this activity is as nonmental brain activity that is a "determinant" of the action. But another way is to view it as the neural realization of a nonconscious distal intention that precedes a *seemingly* spontaneous action. And this interpretation would further support the reply being considered here.

In many cases, however, spontaneous actions occur in response to circumstances that are tokened at or very near the time that the action occurs, such that there is no possibility that a nonconscious distal intention of the kind that Soon, Brass, Heinze, Haynes (2008) may have found, and which may be operative in some cases of seemingly spontaneous action, is formed in advance.

Still, one may challenge the argument on different grounds, namely that it does not give any clear reason for viewing spontaneous, "impulsive" actions as intentional. Perhaps these actions are guided exclusively at the level of sensorimotor control, and do not involve any guidance from intentions. It is not clear what would be missing from such an explanation of these actions, other than their alleged intentional character, which we do not have any independent reason to attribute to them on the basis of this argument.

Though I do not think the argument from indeterminacy and the argument from spontaneous action establish the need to posit executive intentions, I do think they hit on the precise question that a full defense of executive intentions must address: What feature or features of intentional action do executive intentions explain that distal

intentions and motoric states, taken together, do not? I will not be able to launch an exhaustive case here, but I will make some remarks to start addressing this question.

First, to get the ball rolling, let us note that intentions of the distal variety, it is often said, pave the road to hell. This gets at the folk psychological observation that one fails to act on many distal intentions that one forms. There could be a number of reasons for this. Sometimes, for example, one abandons one's intention as circumstances become different and it is no longer appropriate. Or one might change one's mind about what one wants to do. But, among those distal intentions that are not abandoned, why are some acted upon and some not? In order to answer this question, it is useful to posit an additional type of intention that constitutes or results from a *decision* to act on a particular distal intention. These are executive intentions.

But once an agent has decided to act on a distal intention by way of an executive intention, as I am suggesting happens, her executive intention must anchor the action described by the distal intention in the agent's present perceptual context. This *situational anchoring* requires that the agent form an executive intention that incorporates a plan for performing the action at present, which is sensitive to the agent's reasons, beliefs, and desires, and also triggers the appropriate sensorimotor control functions.

In addition, when an agent forms an executive intention, she is not only deciding to act on a distal intention, she is deciding to act in the immediate *present.* This means forming an action plan that specifies the spatiotemporal properties of the action to be

performed—simply put, that it occur here and now.[46] Executive intentions thus provide *spatiotemporal anchoring* of the action.[47] Distal intentions plainly cannot play this role.

I have provided the contours here of a defense of executive intentions—I leave the remaining details for future work. I hope to have provided enough of a foundation, however, that the theoretical merits of positing such intentions, in addition to distal intentions and motoric states, are clear. They help us to explain why certain distal intentions are acted upon, and others are not, as well as to understand how an action is situationally and spatiotemporally anchored within the context in which an agent acts. I move on now to fleshing out some further details regarding their more specific properties.

§3 Filling in the Picture

Though the foregoing provides some insight as to why one might need to posit executive intentions in one's action theory, many questions remain regarding the nature of executive intentions. They are intentional states, yet I have not yet made any remarks concerning their mental attitude and the typical form that their content takes. I turn now to tackling these two issues.

§3.1 The Content of Executive Intentions

[46] More precisely, as I will go on to note in §3, the executive intention need only specify that the action occur now, since that will, by default, determine that it occurs wherever the agent is presently located. The reverse does not hold—if an agent decides to act where she is presently located, she need not have decided to act at that moment.

[47] See also Pacherie (2006, 2008) who defends a *theory of dynamic intentions* on which she distinguishes between distal intentions, proximal intentions, and motor intentions, as I have done here. The notions of situational and temporal anchoring I appeal to here are borrowed from her rich discussion in these papers.

Executive intentions, being types of intention, always represent actions. Executive intentions are always intentions *to do* something. An initial question that arises here is whether executive intentions must, in addition to representing a particular action to be performed, also represent themselves as causing the action. This is sometimes called the *self-referentiality thesis* (e.g., Mele, 1992, p. 197), which is, put more precisely, the thesis that an intention to A is an intention that one As by way of this intention. In other words, the self-referentiality thesis holds that the intention represents itself as the cause of the action it specifies.

Why might one hold this view? One appealing reason is that it is part of the satisfaction or truth conditions of an intention that it causes the action it represents in *some* way (Searle, 1983). Consider a revised version of Davidson's (1973/2001c) climber example. Suppose that the climber forms an executive intention to loosen his hold, but before his intention causes him to do so, he hears a loud noise, it startles him, and as a result, he loosens his hold. In this case, though he does loosen his hold and this is the action his executive intention represents, many would deny that the intention has been satisfied. What must be the case, in addition to the action it represents simply occurring, is that the action be caused *by that very intention*. But here it is not the cause, even indirectly, of the climber loosening his hold.

Perhaps someone will object here that this case does not lend support to the self-referentiality thesis, since the climber's loosening his grip is arguably not an action at all. It is caused by a startle reflex, rather than any intention, and so is better understood as an involuntary movement than an action. And perhaps this feature of the situation, one might urge, is what leads one to view the climber's intention as not being satisfied. But

one can easily address this worry by coming up with a revised case in which an agent forms an intention to A, abandons that intention, and later decides, on the basis of a new intention to A. For example, suppose I intend to go to a concert this evening, but realize I have made plans to see a friend, and so abandon my intention. Upon speaking with my friend, she tells me that she has just been given tickets to this very same concert, and I end up going as part of my plans with her. My going to the concert is certainly an action of mine, but it is not caused by my original intention to go to the concert. This type of case may suffice to convince some that an agent's intention, whether distal or executive, must have as part of its content that the action is the result of that very intention.

This is too quick, though. First, as others have observed (e.g., Pacherie, 2000) if the intention is not satisfied in the type of case described, it need not be because the content of the intention specifies that the intention itself causes the relevant action. It could instead be a feature of the mental attitude of the intention—perhaps intentions are just the kinds of states that must cause the things that they represent in order to be satisfied. Searle, a proponent of the self-referentiality thesis, seems to miss this point when he offers as support an analogy with verbal orders. He writes:

> Suppose I order you to leave the room. And suppose you respond by saying "I am going to leave the room, but not because you ordered me to, I was just about to leave the room anyhow. But I would not have left the room because you ordered me to." If you then leave the room, have you *obeyed my order*? Well, you certainly didn't *disobey* the order, but there is a sense in which you did not obey it either, because the order did not function as a reason for what you did. […] But what this illustrates is that the content of my order is not simply that you leave the

> room, but that you leave the room by way of obeying *this order*; that is, the logical form of the order is not simply: I order you (that I leave the room) but rather it is causally self-referential in the form: I order you (that you leave the room by way of obeying this order). (p. 86)

But far from this analogy serving as support for the self-referentiality thesis, it actually illustrates why it is mistaken. For, though we would indeed deny that one who did not leave the room because she was ordered to, but left of her own accord, was following an order, we also *never* order someone to leave the room by way of saying, "I order you to leave the room by way of obeying this order." So, this case suggests that it is not something in the content of the order that explains why this does not count as following it, but something about ordering, for example, that for an order to be followed it must be the cause of the action it commands.

One may resist this line of reasoning, pointing out that Searle himself takes care to specify that it is the "logical form" of the order that is self-referential. And perhaps one does not read logical form off of the verbal expression of a command, or a mental state, for that matter, but from its satisfaction or truth conditions. Verbal expressions give us rough approximations as to the content of a mental state, one might urge, but they are not themselves the best guides to that content. For this, we must look to satisfaction or truth conditions. And here, doing so suggests that the content of orders, and of intentions, is self-referential. Or so one might argue.

This is a fair point, but it does not help the self-referentiality thesis. For, upon closer examination, we have been given no *reason* to think, in the cases I have presented, that the order to leave the room and the intention to go to the movies are not satisfied.

Perhaps they are not satisfied in the usual way, but why deny that they are satisfied nonetheless? One might say, to distinguish these kinds of cases from ones in which the intention or speech act causes the satisfaction conditions to obtain, that the intention is not *operative* here. But it is not at all clear that it is not satisfied. There seems to be no good reason, therefore, to hold the self-referentiality thesis. An intention to A need not describe A-ing as being the result of that very intention.

But what *is* the content of executive intention? One candidate view, defended recently by McDowell (2010), is that the content of an executive intention is 'I am A-ing'.[48] McDowell takes on board the reasonable view that in order to determine the content of a given mental state, one should look to what one would say in verbally expressing that mental state. So, for example, McDowell holds that an executive intention to raise one's hand has the content 'I am raising my hand', since that is what he takes to be the content of the speech act that expresses that intention. It does not matter, for his purposes, that oftentimes we do not express our executive intentions verbally; what is relevant here is what one *would* say were one to do so.

In endorsing this view, McDowell argues against a competing position, defended by Sellars (1976), that the content of executive intention takes the form 'I shall A now'. McDowell maintains that when one utters, for example, "I shall raise my hand now," the word 'now' signals the *onset* of an executive intention. But, he urges, hand-raisings, like all of our actions, take some time—they are not instantaneous events—so the executive intention is "in action" *throughout* the time it takes for the hand-raising to be completed,

[48] This view is actually inspired by Searle (1983). As such, McDowell (2010) prefers to use the term 'intentions-in-action' to describe executive intentions. But as I said at the outset of this chapter, for present purposes, I will treat these as terminological variants.

and not *only* at the time that one utters 'now'. McDowell concludes that his position is to be preferred over Sellars' position:

> This suggests that the appropriate form for expressing a[n] [executive intention] might be exemplified not by 'I shall raise my hand now', which one might say as one starts to raise one's hand, but by 'I am raising my hand', which one can say at any time during the relevant hand-raising. (p. 416)

McDowell is correct in rejecting Sellars' proposal that the content of an executive intention is 'I shall A now', but not for the reason that he gives. The problem with Sellars' characterization of the content of an executive intention is not that it is not appropriately sensitive to the dynamic nature of action, but that it ignores the distinction between *expressing* and *reporting* intentional states, which Sellars elsewhere takes care to note (see Sellars, 1964; Rosenthal, 2005, p. 266). If one verbally expresses an intentional state, the content of one's speech act matches the content of the intentional state being expressed. For example, if I say that it is raining, my speech act expresses the thought that it is raining, and exhibits the same content as that thought. In addition, the illocutionary force of the speech act corresponds to the mental attitude of the intentional state that it expresses. For example, the assertoric illocutionary force of the utterance that it is raining, corresponds to the assertoric mental attitude of the thought that it expresses (see Vendler, 1972).

When I *report* an intentional state, on the other hand, my speech act does not share the same content as the intentional state that I am reporting on, nor need its illocutionary force correspond to the mental attitude of that state. For example, if I report my hope that it is raining, I might say, "I hope it is raining." And here the content of my

speech act does not match the content of my hope, nor does its assertoric illocutionary force correspond to the mental attitude of hoping, which is plainly not assertoric. In addition, *reports* of intentional states attribute those intentional states to an individual, whereas *expressions* of those mental states do not.

With this distinction between expressing and reporting intentional states in hand, we can see how it is that Sellars' (1976) characterization of the content of executive intention goes astray. Sellars stipulates that he will adopt a convention where 'shall' is used to "express" executive intention. So an executive intention to raise my arm is expressed, on Sellars' view, by saying, "I shall raise my arm now." But he also says, in the same breath, that he will treat 'shall' as the mental attitude associated with executive intention. If so, however, then this verb has no place in a verbal *expression* of an executive intention, but rather in *reports* of executive intention. The verbal expression of a mental state, as illustrated above, does not include in its content the mental attitude of the state it expresses—this is merely *reflected* in the illocutionary force of the speech act.

I will say more about this point in the following section, but for now, if we adjust Sellars' view to be sensitive to the distinction between expressing and reporting, the correct content of an executive intention, using the 'shall' construction, is not 'I shall A now', but 'I A now'. Given this amendment, we must ask whether McDowell's argument against Sellars' original, and I have argued mistaken, proposal has any force against this revised version. I will argue that it does not.

A crucial premise in McDowell's argument against the view that the content of executive intention is 'I shall A now' is that the executive intention remains active throughout the action that it causes. Indeed, theorists commonly attribute to executive

intentions not only the role of initiating actions, but of sustaining and guiding them (e.g., Mele & Moser, 1994). After all, they urge, if one abandons an executive intention halfway through an action, one ceases performing the action, and this suggests that executive intentions not only initiate, but help sustain an action through to its completion.

But even still, this is not sufficient to show that the content of an executive intention is 'I am A-ing' rather than 'I A now', as McDowell claims. This is because, while one can truthfully say 'I am A-ing' any time during an act of A-ing, one can also truthfully say 'I A now' at any time during an act of A-ing. As I am raising my hand, it is both true that I am raising my hand, and true that I raise my hand now. McDowell restricts the referent of 'now' to the onset of the action, but it is not clear why he does this. Sellars' construal of the content of an executive intention does not rule out an executive intention's being active for the duration of an action, as McDowell seems to hold.

There is an important asymmetry lurking, however, with respect to the appropriateness of verbally expressing the content of an executive intention that we have been considering. While one can truthfully say 'I A now' at the moment that an executive intention initiating an action is tokened, one cannot say 'I am A-ing', for at that point, no action is yet underway. Since executive intentions initiate intentional action by being the most proximate mental states to cause it, this point decisively rules against the view that their content is 'I am A-ing' and in favor of the view that it is 'I A now'.

Still, one sometimes *does* say that one is A-ing while one is performing an action. This is especially clear in cases where an action's duration is quite long, for example, walking several blocks to the neighborhood park. If one is asked what one is doing during

this time, one might very well say "I am walking to the park." But if one does not express an executive intention in such cases, as I have just argued, then what mental state *does* one express? The answer is that one expresses a thought to the effect that one is A-ing, or walking to the park. I will say much more about the relation between executive intentions to A now and thoughts to the effect that one is A-ing in the forthcoming sections. But for now, I turn to make some remarks on another aspect of the content of executive intention.

I have been arguing that the content of executive intention takes the form 'I A now'. But I have not yet said anything about what has been called the "unavoidably first-personal" (Castañeda, 1972) character of this content. Executive intentions, like their distal counterparts, always represent an action that one is, oneself, to perform. They cannot be about actions that other agents are to perform. One might colloquially say, for example, that one intends for his friend to come to dinner, but this is a loose way of saying that one intends for one, oneself, to bring it about that one's friend comes to dinner. This type of case does not entail that intentions are sometimes intentions for someone else, and *not oneself*, to do something.

Some might object, however, that intentions do not make explicit reference to oneself in their content, despite its being the case that one is always, oneself, the agent of the action that they represent. Rather, one might insist that the content of an executive intention is simply an action—that an executive intention is simply, as Baier (1970) put it, "an intention to do." The fact that executive intentions do not directly cause actions that any *other* agent performs is, on this view, an artifact of their being exclusively hooked up to the motor system of the individual that holds an executive attitude towards whatever action is represented. It is not a built-in feature of their content.

But this view cannot account for the types of relations that executive intentions enter into with other mental states. Importantly, executive intentions, and intentions more generally, stand in *inferential* relations to other mental states, and this would not be the case if their content were simply an action, since no inferences can be made on the basis of such content. Nothing can be inferred, for example, from the content 'to walk', but inferences *can* be drawn from the content 'I walk', or 'I will walk', such as the inference that one is not running, or that one is planning to engage in activity requiring leg movement.

§3.2 'Willing' as the Mental Attitude of Executive Intention

I propose that to have an executive intention to A is to will to A.[49] How does willing as a mental attitude fare with respect to the distinction between expressing and reporting? On the present account, when one says, "I will A now," one is thereby reporting one's executive intention. One may resist this claim, since the full content of the executive intention, as I argued in the previous section, is "I A now," which implies that the report should be, "I will that I A now." But when verbally reporting on our own mental states it is common to elide over features of the content that add no informational value. Since one is already attributing the intention to oneself by way of the report, and

[49] When Sellars (1976) discusses this option, he speculates that it will strike a number of people as "contrived to save a theory which urgently needs a verb to support the verbal noun 'volition'" (p. 47). He goes on, however, to offer by way of reassurance the claim that first-person expressions of intention do make use of the verb 'to will', as in 'I will go to the park'. While I am sympathetic to the spirit of his point, the details strike me as incorrect, as I pointed out above—the verb 'willing' does not occur in first-person *expressions* of intention, but rather in first-person *reports* of intention. And this applies to third-person ascriptions as well—in describing another agent's present actions, one often says things like 'S will now A' to describe them, again implicating the verb 'to will' in intention more generally.

one's audience is aware that intentions in all cases specify an action that one is, oneself, to perform, one may naturally skip over the additional reference to oneself in the content of the intention. Similarly, when reporting a desire for a vacation, one may simply say, "I want a vacation," which is elliptical for "I want that I have a vacation." In the case of an executive intention, saying, "I will A now," is elliptical for "I will that I A now," which more accurately reflects the content of the state being reported upon.

It is true that, unlike in the case of belief and other mental states, we rarely give verbal expression to our executive intentions. It is common to make utterances like, "it is raining," but very rarely do we say things like, "I walk now." There is a ready explanation for this asymmetry: unlike most other mental states, executive intentions find their most natural expression in nonverbal behavior, that is, in bodily movement. This is because they are, among an agent's many intentional states at the time of intentional action, the most proximate causes, unlike beliefs and desires, which play more distal roles in bringing it about. Given that their most natural expression is in nonverbal behavior, in many cases the information that would be provided to a listener by verbally expressing an executive intention is redundant—it is already readily observable on the basis of one's behavior.

In some cases, though, it may be useful to verbally express one's executive intention in order to indicate aspects of one's behavior that are intentional. It may be clear to an observer that one is walking somewhere, for example, but not clear, unless one verbally expresses one's executive intention, that one is going to the park. Even here, though, it will often be more natural to express not the executive intention to go to the park itself, but the thought that it results in to the effect that one is going to the park.

More generally, this is the difference discussed previously between saying, 'I A now' and 'I am A-ing'—the former being an expression of an executive intention, and the latter being an expression of a thought. Once one has reached maturity in motor development, executive intentions, being reliable causes of the actions they represent, under most conditions lead seamlessly to a thought that one is performing the action in question. This means that, often, rather than expressing the executive intention to A, one can safely, that is, without error, express the thought that one is A-ing.

§4 Executive Intentions and Agentive Awareness

Though details remain to be filled in—I leave that for future work—the foregoing sketch suffices to provide us with a working model of executive intentions. How does all this relate back to agentive awareness? In the remainder of this chapter I will defend an intention-based account of agentive awareness. I distinguish in this section between two different forms such an account might take.

§4.1 The Constitutive Account

One version of the intention-based approach to agentive awareness is what I will call the *Constitutive Account*. This account takes agentive awareness to be constitutively tied to executive intentions. On this view, executive intentions themselves, like perceptions, are states capable of making an individual aware of things. Call this type of awareness *volitional awareness*. And since executive intentions represent oneself as performing particular actions, executive intentions thereby make one aware of oneself as such. We may characterize the central claim of this account very simply as follows: Agentive awareness is volitional awareness. In other words, what it is to be aware of

oneself as performing an action is to represent oneself as performing an action by way of an executive intention—the very states that initiate and help to guide intentional action.[50]

Such an account may seem doomed from the outset. After all, executive intentions do not seem to be the kinds of mental states that make one aware of anything at all, let alone one's actions. Perceptions are paradigmatic states that make us aware of objects and states of affairs. I can perceive, for example, that an apple is red, or that my legs are crossed. And, arguably, thoughts sometimes make us aware of things as well—in particular, thoughts about things as being present. But perceptions and thoughts share in common a feature that many view as the feature in virtue of which they make us aware of things: they are assertoric states—they represent the world *as* being a certain way. My perception that my legs are crossed represents my legs as being crossed. And my thought that I am in my apartment represents me as being in my apartment. This feature of perceptions and thoughts, however, appears not to be shared by mental states like hopes, wishes, and desires. I cannot be aware of something by having a hope about it, or having a wish about it, at least not unless I have an assertoric state like a perception or a thought about it *first.* And executive intentions, one may urge, belong in the same family as hopes, wishes, and desires, so they, too, are not states in virtue of which we become aware of things.

Another way to make this point is by appeal to the notion of *direction of fit*. Following Anscombe (1966), and later Searle (1983), a distinction is commonly made

[50] This account is Ancombean in spirit, given that Anscombe (1966) defended the view that we can have "knowledge without observation" of what we do intentionally. There are important differences between Anscombe's treatment and my own—most importantly perhaps, that I do not take epistemic considerations as my focus as she does, and she did not, as I the account I am sketching here does, treat the states in virtue of which we have such knowledge as executive intentions.

between mental states with a *world-to-mind* direction of fit and mental states with a *mind-to-world* direction of fit. For ease of exposition, and following Humberstone (1992), Bayne (2011a), and Pacherie (forthcoming), I will call mental states with a world-to-mind direction of fit *telic* states and states with a mind-to-world direction of fit *thetic* states.

Telic states dispose or cause one to respond in ways that make their content true, where this includes dispositions to form relevant mental states as well as to engage in appropriate behavior. Desires are paradigmatically telic states. A desire to have a glass of water, for example, disposes me to go get myself a glass of water. Thetic states, on the other hand, dispose or cause one to respond as if their content is true. Perceptions and thoughts are paradigmatic thetic states. If I perceive or think that it's raining, for example, then I will be disposed to respond as if it's raining by, for example, taking my umbrella when I go outside.

The problem with executive intentions, one might complain, is that they are best understood as telic states rather than thetic states. After all, a central function of executive intentions is to cause one to act in the ways that their content specifies, that is, to act in ways that makes their content true. And if awareness is granted by thetic states only, then executive intentions are not themselves the mental states in virtue of which we are agentively aware. For this we must appeal to a *bona fide* thetic state, such as a perception or a thought.

The proponent of the constitutive account might respond to this worry by insisting that it rests on a false premise. The reasoning I have presented is that since executive intentions are telic states, they are not thetic states, thus smuggling in the assumption that mental states are either telic or thetic, but not both. It may be, however, that executive

intentions have a *dual* direction of fit—that they are properly understood as being both telic and thetic at once.[51] And if they are both telic and thetic states, then they can make one agentively aware in virtue of their thetic properties, and bring about actions in virtue of their telic properties.

The idea of a mental state with a dual direction of fit may at first seem mysterious. But in fact we have a familiar model of this when it comes to certain types of speech acts—and we may in general glean valuable insights about the nature of intentional states from examining the nature of speech acts (cf. Rosenthal, 2005; Searle, 1983; Sellars, 1964). The speech acts I have in mind are what are known in speech act theory as *performative utterances* (Searle, 1989). Examples of performative utterances are: "I order you to leave the room," "I promise I will pay you back," and "I apologize for being late." Such utterances have a peculiar quality; in saying, "I order you to leave the room," not only do I assert of myself that I am ordering you to leave the room, but I also make it the case that I am ordering you to leave the room in performing that very utterance.[52] There seems to be no conflict between these two distinct roles—of asserting

---

[51] Bayne (2011a) explores the related possibility that agentive experiences are pushmi-pullyu representations. The account developed here is different from Bayne's, however, as he sees agentive experiences as dynamic states that have telic structure early on and thetic structure towards the end. My claim, rather, is that agentive states have executive and assertoric structure simultaneously. Moreover, Bayne does not defend the view that agentive awareness is tied specifically to executive intentions, as I am suggesting here. Indeed, he explicitly rejects such a view in Bayne (2011b).

[52] There has, of course, been much debate over how to understand performatives within the philosophy of language. J. L. Austin (1975), a pioneer in this area, held that in normal circumstances, in saying 'I order you to leave the room', I thereby perform the act named by the main verb in my utterance. But, he urged, appearances to the contrary, I do not also make a statement. Others, however, have taken issue with Austin's analysis, and argued that things are just as they seem with performatives, i.e., that they are both statements and performances (Bach, 1975). For purposes of this analogy, I assume this latter interpretation.

something and making that very thing the case—of one and the same utterance.

When it comes to executive intentions, the proponent of the constitutive account may suppose that they work similarly in having a dual direction of fit. An executive intention that I now A both causes one to behave as though one is A-ing, and causes one to behave in ways that make it the case that one is A-ing. Now, presumably, the latter role of executive intentions is one that few would deny—a central way of conceiving of intentions in general is as states that drive one to perform the action that they represent, and things are no different for executive intentions, which, if anything, have a tighter connection with the occurrence of action than their distal siblings. Where further work is needed is in establishing that executive intentions also have a thetic component. And for this, given the characterization of thetic states I offered above, one may argue that they do, in fact, dispose one to respond as though their content is true.[53]

Indeed, as I will go on to elaborate in §4.2, is it often argued that if one intends to A, then one believes that one will A. If this is true, then it seems that intentions meet the condition I have specified for being thetic states. But, supposing it is true, we must ask *why* it is true. Why would it be the case that an intention to A results in the belief that one is A-ing. The most credible answer here is that it is because intentions reliably cause the actions that they represent, and one comes to *learn* that this is the case. In learning that this is the case, one comes to form beliefs on the basis of one's intention to A, that one will A. But this only lends intentions a *quasi-thetic* character, for were it not for this

[53] The general spirit of this proposal is not without precedent. There is already a tradition within action theory known as *cognitivism*, which seeks to characterize intentions as thetic states (Setiya, 2007; Velleman, 1989). More specifically, this view attempts to assimilate intention to a type of belief, but for our purposes, we need only focus on the more general claim that intentions are thetic, regardless of whether they are types of belief in particular.

reliable connection—were it to be the case that, due to some impairment, one was unable to ever act on one's executive intentions—this connection would not hold. In other words, it is not a feature of the mental attitude of willing that it is thetic, but a feature of the typically reliable connection that intentions have with actions.

To see this point more clearly, consider by analogy the case of an omnipotent being whose desires are always immediately satisfied. Confronted with such a being, we might come to view her desires as guides to the way the world is—as the basis for the beliefs that we form about the world. But this would not make them any less desire-like, or transform them into types of beliefs. At best, they would have what I am describing as the quasi-thetic status that executive intentions typically have. We have no good reason to view executive intentions as having a dual direction of fit.

Still, even if I am wrong about that, and executive intentions *do* have a dual direction of fit, there remain obstacles to viewing them as the states in virtue of which we are agentively aware. On the constitutive account, if one forms an executive intention to A, then one will, in virtue of this very executive intention, be aware of oneself as A-ing. But one may challenge this implication of the view since, at first blush, there are cases in which one is doubtful of the success of one's action, despite forming an executive intention to perform it. Call this the *objection from uncertainty*.

Davidson's (1978/2001d) classic carbon copier example will serve to illustrate this worry. In Davidson's example, an agent presses down on a stack of papers with the aim of making ten carbon copies. But as he does so, he doubts that he is actually succeeding in making ten carbon copies. If the following further claims about the scenario are true, then this case presents a problem for the constitutive view: (i) the

carbon copier forms an executive intention to make ten carbon copies, and (ii) one cannot simultaneously be aware of oneself as A-ing and doubt that one is A-ing. If so, then despite forming an executive intention to make ten carbon copies, the carbon copier is not thereby aware of himself as making ten carbon copies, since he is doubts that he is doing so. And if that is true, then executive intentions cannot by themselves be the states in virtue of which one is agentively aware.

One might reply to the objection from uncertainty by denying that the carbon copier forms an executive intention to make ten carbon copies. This move gains some plausibility once we question the source of the carbon copier's uncertainty or doubt that he is succeeding in making ten carbon copies when he presses down on the stack of papers. Here, and in cases like this one, the uncertainty is likely anchored in a doubt that he will succeed in making ten carbon copies that is present *prior* to committing to the course of action he is taking. But if so, then this doubt affects the subsequent formation of his executive intention. If one doubts that one will be able to A, then rather than form an executive intention to A, in such a case, one will arguably instead form an executive intention to *try* to A. And such an intention would not create any trouble for the constitutive account, since there is no problem with the carbon copier being aware of himself as *trying* to A in virtue of his executive intention to try to A.

But one might insist here that the carbon copier does not *merely* intend to try to make ten carbon copies, but intends to *make* ten carbon copies. After all, if he presses down on the stack and fails to make ten carbon copies, it would be odd to say that he nonetheless succeeded in his goal, though it seems we would be forced to say just that if he were merely trying to make ten carbon copies. This response ignores the

consideration I have offered that one's intentions are typically based on what one believes that one is able to do. Even if we allow the response, however, there are other ways that the agent might adjust his executive intention in light of uncertainty with respect to successfully performing the action. He might, for example, form an executive intention to do something that is a means to A-ing, but is not itself A-ing, in the hopes that securing the means to A-ing might result in A-ing. For example, he might form an executive intention to press down hard on the stack of papers—with the further intention of making ten carbon copies. Doubtful of success with respect to some action, one does not simply proceed with one's plans involving that action unfettered; the executive intentions one forms are sensitive to what one takes one's abilities and prospects of success to be. And if the carbon copier merely forms an executive intention to press down hard on the stack of papers, with the hope that this will result in making ten carbon copies, then this case actually seems to deliver a verdict compatible with the constitutive account. The carbon copier is aware of himself as pressing down hard on the stack of papers, rather than aware of himself as making ten carbon copies.

In addition to this reply, the proponent of the constitutive account has reason to deny the further premise the objection from uncertainty relies on, that is, that one cannot simultaneously be aware of oneself as doing A and doubt that one is doing A. To see why, consider an analogous scenario in which a conflict holds between one's perceptual awareness and one's beliefs. For example, consider the so-called motion after-effect (or "waterfall") illusion, which occurs after staring at a moving stimulus for a suitable amount of time and then fixating on a stationary stimulus (for review, see Mather, Pavan, Campana, & Casco, 2008). The stationary stimulus will appear to move in the opposite

direction of the moving stimulus. Still, one who is undergoing the illusion, but has been told in advance about it, will believe that the stimulus remains stationary. So in this case, one has perceptual awareness that the stimulus is moving, and a simultaneous belief with conflicting content, namely that the stimulus is stationary. Applying this to the carbon copier case, it might similarly be true that the agent has volitional awareness that he is making ten carbon copies, and a simultaneous conflicting doubt or uncertainty that this is the case. Such a dissonant psychological global state would arguably last only briefly. Persisting uncertainty about the success of the carbon copier's endeavor would move him to form a new action plan, and thereby a new executive intention. But, just as in the waterfall illusion, there may be no problem with a case in which his volitional agentive awareness is in tension with his belief about what he is doing, at least for a short time. I conclude that the objection from uncertainty does not go through.

Another objection to the constitutive account of agentive awareness comes from the case of skilled action. Call this the *objection from skilled action*. The argument on which the objection is based may be motivated by way of the following remark by Sellars (1976):

> … if one wills to play a melody, there need be no volition, with respect to each note, to sound that note. It is those who are learning to play the melody who are in this position, as was the fabled centipede who fell into confusion when he was asked how he walked. (p. 50 – 51)

In models of skill acquisition, it is commonplace to posit that in the early stages of learning, novices follow an explicit set of rules for the novel action (e.g., Fitts & Posner, 1967). It is arguable that when one learns a new skill, then, such as playing a quick

arpeggio on the piano, doing a backhand serve in tennis, or even tying one's shoelaces, one forms an executive intention for each stage of the action that corresponds to the rules that one is carefully following each step of the way. These executive intentions are, in a sense, ways of internalizing these rules. But once one has become proficient at the skill, it seems, as Sellars suggests, that one not only stops following explicit rules, but also stops forming corresponding executive intentions each step of the way. Perhaps one need only form an executive intention to perform the action at the outset, and the learned associations at the level of sensorimotor control largely take over the rest of the way. Still, one does not deny authorship of the actions that are not represented in one's general executive intention, so one has agentive awareness over them in the same way that one has agentive awareness over the action as a whole. But whatever agentive awareness one has of these actions, then, cannot be had in virtue of an executive intention. This would be a problem for the constitutive view.

This argument relies on some controversial premises. For one, there is the central claim that after the learning phase, as one progresses from novice to expert, one no longer forms executive intentions for each step of the skilled action. This is to take on board a position which is in the spirit of a view recently defended by Dreyfus (2005, 2007a, 2007b; see also McDowell, 2007; Montero & Evans, 2011). Dreyfus argues that people no longer think about what they are doing when they have achieved expertise with respect to a certain skill—instead acting in an automatic, unreflective, immersed way. According to Dreyfus, this allows us to understand the impressively high-level of flexibility and responsiveness to the details of their situation of action that experts display. Experts would not be able to exhibit this kind of rapid reactivity, Dreyfus urges,

if they were thinking, or forming executive intentions for that matter, prior to their every move—this would just slow them down.

It is important to note that the objection from skilled action, as well as Dreyfus's view, is committed to there being no executive intentions involved at the level of skilled action *whether conscious or not*. Skilled action, on this view, is not a matter of going from conscious to nonconscious intentions, it is a matter of doing away with thinking about the action altogether, where this includes forming executive intentions. But the justification that Dreyfus cites in support of expert action involving no thought—and presumably the one that would be cited by the proponent of the objection from skilled action with respect to intention—only seems to apply to the case of conscious thought—the point is that individuals would be slowed down. And though there may be reason to think that *consciousness* slows things down, this is not the same as *thinking* or *intending* slowing things down.[54] So the objection requires further support here.

In addition, the objection from skilled action relies on the false premise that if one does not deny authorship of an action, then one is agentively aware of it. As mentioned in Chapter 4, it could be instead that one is neither agentively aware of them, nor aware of them as alien—indeed, this is what would be predicted if one were simply *not* aware of many of one's skilled actions as one is performing them. And this seems to capture how many experts report on their own skilled actions. Consider, for example, the following

---

[54] Dreyfus is here endorsing a more general version of what Montero (2010) has called "the Maxim", which is the more specific view that attention to bodily movement typically interferes with expert level performance. Montero convincingly argues that there is reason to be skeptical of the Maxim as it applies to expertise, since part of what it is to be an expert is to continually improve on one's performance by way of attending to one's technique. Still, the more general claim that *consciousness* tends to interfere with skilled bodily movement, which is the one at issue here, is untouched by Montero's critique.

quotation from Larry Bird, the NBA basketball player, who once said: "[A lot of the] things I do on the court are just reactions to situations . . . A lot of times, I've passed the basketball and not realized I've passed it until a moment or so later" (as reported in Dreyfus & Dreyfus, 1999, p. 112). One may come to be aware of what one has done by way of consciously observing one's bodily motion, or the consequences of action. But, as these are cases of forming agentive judgments on the basis of conscious observation and inference, they are not cases of agentive awareness, which, as I have stressed throughout this dissertation, is *subjectively unmediated* awareness of oneself performing an action *as one is performing it.* The objection from skilled action does not go through.

Another objection that may be raised against the constitutive account is the *objection from action side-effects*. According to this objection, one may sometimes be agentively aware of actions that one does not intend to perform, but that one nonetheless expects one will perform as a side-effect or consequence of whatever action one does intend to perform. To take an example adapted from Paul (2009), one may form the executive intention to park one's car in a tight spot, while believing that if one succeeds in doing so, one will scratch one's vehicle. Of course, one does not intend, or have an executive intention, to scratch one's car. Still, one is agentively aware of oneself as scratching one's car as one is parking it. But if so, then this agentive awareness is not had in virtue of any executive intention, since one does not intend to scratch one's car.

Once again, this objection does not succeed because this is not a case in which one has subjectively unmediated awareness of one's action. One's awareness that one is scratching one's car is in this case based on the conscious observation that one is parking one's car, and a conscious inference to the effect that if one is parking one's car, one is

also scratching it. But as I said in response to the last objection, these conscious inference and observation based agentive judgments are not the kinds with which we are concerned.

The constitutive account of agentive awareness thus seems to get around several initial objections that may be raised against it. Still, there is at least one empirical result that I take to rule against the constitutive account. Recall Libet, Gleason, Wright, & Pearl's (1983) classic experiment in which they asked participants to perform a simple action, such as a flexing of the wrist, and to report the time at which they were aware of a decision to act (W-judgment) as well as the time at which they were aware of actually acting (M-judgment). Now the W-judgment is, in effect, a report in the timing of one's awareness of an executive intention. Executive intentions, as I have characterized them, are simply decisions to act now. And the M-judgment is, in effect, a report on the timing of one's agentive awareness, since it is a report on when one was aware of oneself as acting. But the time reported in the M-judgment and the time reported in the W-judgment, recall, were separated by about 120 ms. This presents a serious problem for the constitutive view, since it suggests that one's awareness of acting is not constituted by one's executive intention—if the judgments are reliable, then the executive intention occurs about one tenth of a second *earlier* than the subsequent awareness of acting.

I propose that, in light of this result, the state in virtue of which one is aware of oneself as acting is not, after all, an executive intention, but is instead a thought to the effect that one is A-ing, which is *based* on one's executive intention to A. Call this the *inferential account of agentive awareness*. Below I sketch further details of the account,

and show how it meets all the desiderata of a theory of agentive awareness laid out in Chapter 1.

§4.2 The Inferential Account of Agentive Awareness

On the inferential account of agentive awareness, one is aware of oneself as A-ing by way of a thought that one is A-ing. One's thought that one is A-ing is, in turn, based on one's executive intention to A. In order to get a better handle on the account, therefore, we must start by exploring this alleged connection between an executive intention to A and a thought that one is A-ing.

In both commonsense and more formal theorizing (e.g., Grice, 1972; Harman, 1976), many support what Mele (1992) calls a "strong belief requirement" for intention. On such a requirement, an agent S intends to A only if S believes that he will A.[55] The motivation for a strong belief requirement is well-illustrated in cases like the following, offered by Grice (1972, p.4):

> X: I intend to go to that concert on Tuesday.
>
> Y: You will enjoy that.
>
> X: I may not be there.
>
> Y: I am afraid I don't understand.

Given X's uncertainty about whether he will, in fact, go to the concert on Tuesday, Grice urges, one should expect him to add qualifications, such as, 'I intend to go to that concert on Tuesday *if* I am not working' or 'I intend to go to that concert on Tuesday *if* my cold has passed' in reporting his intention to Y. But given that X does not add such qualifications, X misleads Y in saying that he intends to A, but may not, in fact, A—or so

[55] Or, as Grice (1972) puts it: If S intends to A then "S should be sure that he will in fact do A" (6).

Grice and others have supposed. While this case involves distal intention, the general rule, that one intends to A only if one believes that one will A, could be extrapolated to apply to the case of executive intention, so that one wills to A now only if one believes that one is A-ing.

There is reason to deny that Grice's (1972) case presents support for a strong belief requirement. For instance, Davidson (1978/2001d) convincingly argues that what follows from the case is that there is a sociolinguistic norm governing the reporting of one's intentions, i.e., that when one sincerely reports that one intends to A, one gives license to the speaker to believe that one will A. And since one speaks sincerely, one also gives the speaker license to believe that one believes that one will A. So there is a connection here between sincerely saying that one intends to A, and giving license to a speaker to believe that one believes that one will A. But this does not support the much stronger claim that if one *intends* to A, one believes that one will A.

There are replies that the proponent of the strong belief requirement can give here, but thankfully, I do not need to settle this issue here. For present purposes, all that must be established is that *some* reliable relationship holds between an executive intention to A, and a thought that one is A-ing. And there is good reason to think that an executive intention to A *disposes* one to think that one is A-ing, even if it does not in all cases give rise to such a thought.

Executive intentions to A reliably cause actions of A-ing. This simple observation was an important point in Chapter 4, where I argued that Wegner's (2002) matching model of agentive awareness is not appropriately sensitive to the high degree of reliability that the connection between an intention and the action it causes displays. But

it can be put to work here as well, in explaining how it is that executive intentions can give rise to thoughts in virtue of which an agent is agentively aware.

The causal connection between an executive intention and the action it represents is at its most reliable when A-ing is a bodily movement, such as raising one's arm, or a bending of one's elbow. For in these types of case, the probability of something going wrong, in a healthy, well-functioning agent is very low. It is relatively rare, and it is worth taking a moment to reflect on just how rare, that one forms an executive intention to perform a simple bodily movement that fails to be satisfied.

And presumably, after some time, an agent comes to learn that these reliable connections hold between her executive intentions and the bodily actions they represent. Arguably, the agent does this early on by way of the psychological model that Wegner (2002) describes. She may learn that her executive intentions reliably cause corresponding actions by way of observing a match between what she wills to do and what she actually does do, and inferring on this basis that her executive intention is the cause of her action. Wegner's model is indeed attractive as an explanation of how one comes to form one's early judgments about the causal connection between one's intentions and one's bodily actions.

But once the agent has learned that the causal connection is a highly reliable one, and this is the point that Wegner's model overlooks, she typically no longer needs to wait to observe a match between her executive intention and her bodily action in order to infer that her executive intention causes her action. She can rely instead on the executive intention itself as well as a standing belief, acquired as a result of her learning the relationship between her executive intention and her bodily actions, that she tends to

perform whatever action her executive intention specifies. So if she forms an executive intention to A, given her standing belief that she tends to A whenever she has an executive intention to do so, she is able to infer on that basis that she is A-ing—it is not necessary to spend valuable time confirming by way of perceptual observation that her executive intention successfully causes the bodily action it represents.

And once this inference becomes relatively habitual, it will take on an automatic character. After all, inferring is a mental action, and there is no reason to deny that mental actions can become as automatic as bodily actions, once they become well-practiced. So when the agent forms an executive intention to A, she will also be strongly disposed to form an accompanying thought to the effect that she is A-ing.

On the view that is emerging, the default stance a mature agent takes is that her executive intentions, especially those pertaining to her bodily actions, are successful. The probability that things have gone well typically outweighs the probability that they have not. So when she forms an intention to A, she is also disposed to think that she is A-ing. Sometimes this disposition may not manifest itself. This may happen in cases in which the success of the action is uncertain. I catalogued some of these possible cases in Chapter 4: performing an action in challenging settings (e.g., underwater, in the dark, on a shaky bus); performing an action in a situation where stakes are high (e.g., in a physically dangerous situation, in a socially demanding setting); performing an action when one's regular capacities are impaired (e.g., after temporary paralysis, when one is intoxicated, when one is exhausted). A full inferential theory will need to determine just what the relevant principles are that determine when one's disposition to infer that one is

A-ing is manifested, and these principles may be relative to individual agents and their background beliefs.

There are also situations, of course, when something goes *wrong* with an action after it has already begun. If the agent is aware of evidence that what her body does fails to correspond to her executive intention, she will, of course, no longer think that she is A-ing. But in general, she does not seek confirmation of her default thought that she is A-ing, rather she seeks *dis*confirmation.

I have been focusing on bodily actions here, but the same model applies, *mutatis mutandis*, to the case of actions incorporating non-bodily events. Once a reliable connection is learned between one's executive intentions and *these* types of actions, one no longer needs to monitor one's actions to establish a link between them and one's executive intention. One's executive intention to A will be sufficient on its own to dispose one to think that one is A-ing.

The foregoing is a start, but it is only a sketch of the inferential account. Future research in philosophy and cognitive science will need to spell out in further detail how the inferential account works. But in the meantime, the theory gets off the ground with the help of two pillars of initial support that it enjoys. First, it successfully avoids the problems that each of the other accounts I have reviewed faces. The inferential theory is a non-sensory, high-level, non-matching account of agentive awareness. Second, further support for the inferential account—as I will now argue—derives from how well it performs in meeting the desiderata that I laid out in Chapter 1.

Recall that in Chapter 1 I said that any theory of agentive awareness should successfully explain the following five features of agentive awareness: (i) its self-

referential character, (ii) its subjectively unmediated character, (iii) its phenomenal character, (iv) its representational character, and (v) its absence in disorders of volition. I also said that a theory of agentive awareness should have something to say about cases in which it breaks down. I turn now to evaluating the inferential account on the basis of these criteria.

§5 Evaluating the Inferential Account

§5.1 Self-Referential Character of Agentive Awareness

The inferential account is perhaps at its strongest when it comes to explaining the self-referential character of agentive awareness.[56] For the type of thought in virtue of which one has agentive awareness, on this account, is a first-person thought with the content 'I am A-ing'. Such a thought makes reference to oneself by way of what has been labeled the *essential indexical*. Essentially indexical thoughts and speech acts take the form 'I am F', and in deploying the essential indexical, they ensure that the reference they make to oneself is reference to oneself *as such*. As an illustration, consider John Perry's (1993) classic example, in which I am walking around the supermarket and notice that someone is spilling a trail of sugar. I think to myself that someone is making a mess, not realizing that it is I who am making a mess. After a while, I come to have the thought that *I* am the one that is making a mess. Now, though both my earlier thought that someone is making a mess, and my later thought that I am the one that is making the mess have the same referent, namely myself, it is only the latter thought that makes reference to myself *as such*.

[56] This treatment of the self-awareness provided by first-person agentive thoughts is much influenced by discussions with David Rosenthal, and by his published work on the topic (see Rosenthal 2005, Ch. 13; Rosenthal, 2012).

It is this kind of reference to oneself that is required for agentive awareness. It will not do, in order to be agentively aware, for one to be aware that *someone* is performing a particular action, where that someone is in fact oneself, without one's being aware that it is oneself, *as such*, who is performing it. Agentive awareness requires awareness of oneself *as such* performing an action, and the essential indexical that appears in the content of first-person thoughts is a way of securing such awareness.

But what is it about the essentially indexical self-reference that makes one aware of oneself *as such* by way of first-person thoughts that employ it? If I have a thought that I am A-ing, my thought, by way of the essential indexical, in effect represents as A-ing the thinker of that very thought. It forms a tacit identification between the thinker of the thought and the agent of the action. This is what constitutes self-reference *as such*—that is, self-reference that is independent of any particular description apart from being the bearer of the very thought about oneself. Going back to Perry's supermarket case, what makes me recognize that *I* am the one spilling the sugar, in the relevant way, is that I come to realize that the thinker of the thought that somebody is spilling the sugar is the very person that is spilling the sugar.

When it comes to agentive awareness, the essentially indexical self-reference secured in first-person thoughts to the effect that I am A-ing provides a tidy explanation of the strong intuitive sense of *self as source* that so many theorists view as being a feature agentive awareness (Bayne, 2006; Horgan, Tienson, & Graham, 2003; Horgan, 2012). The self in question is simply the very thinker of the thought that one is A-ing.

§5.2 Phenomenal Character of Agentive Awareness

Recall from Chapter 1 that many find it obvious that just as there is a phenomenology to visual perception, pain, emotion, and other mental phenomena, there is a distinctive phenomenology of agency. But thoughts, many urge, do not have any characteristic phenomenology. So how can a theory on which an individual is made agentively aware by having a thought about one's action account for the phenomenology of agency?

A theory of agentive awareness has two options when it comes to successfully explaining the seeming phenomenal character of this type of self-awareness. One is to explain how it is that the states and processes appealed to in the theory could give rise to such phenomenology, and another is to explain why it is that it has *seemed* to so many that there is a phenomenology of agency—that is, to provide an error theory. Both approaches have promise when applied to the inferential account.

First, there is a growing literature on cognitive phenomenology, with many theorists arguing that thoughts have phenomenal character no less than sensations, perceptions, emotions, and the like (for recent contributions to this debate, see Bayne and Montague, 2012). Indeed, Horgan (2012) uses agentive phenomenology as a stepping stone to argue for the existence of cognitive phenomenology more generally.

In addition, I have just described how agentive thoughts might give rise, by way of the essential indexical, to what many take to be a central feature of the phenomenology of agency, namely the sense of self as source. The inferential theorist need not despair, therefore, that such an account must do away with the phenomenology of agency.

The second option, of providing an error theory for why people have thought there is a phenomenology of agency, has merit as well. One reason that people might

think that there is a phenomenology of agency, is that there is a phenomenology of being *alienated* from one's actions, and the phenomenology of agency is often picked out in a contrastive way as being *not* like that. When something goes wrong with one's action, one often has a sharp and sudden awareness that this is the case. This is plausibly an affective reaction, a distress signal of sorts that indicates to one that something is amiss. In describing the phenomenology of agency, people often point to these types of experiences first, and say that by contrast, when actions are going smoothly there is a corresponding phenomenology. But it may very well be that there is merely a phenomenology of alienation, and no phenomenology of agency at all. Indeed, Prinz (2012) has recently endorsed such a deflationary view. I will not settle here which of these two options is the one to pursue—I merely wish to show that an inferential account of agentive awareness has the resources to explain the phenomenology of agency, either by appeal to cognitive phenomenology or a deflationary account.

§5.3 Subjectively Unmediated Character of Agentive Awareness

One worry that may confront one in evaluating the inferential account on the basis of the desiderata I laid out in Chapter 1 is that it fails to explain the subjectively unmediated character of agentive awareness. After all, on the inferential account, agentive awareness is provided by thoughts that are inferred on the basis of our executive intentions and background beliefs about the efficacy of those intentions. But such inferences would seem to threaten the inferential account's ability to explain the immediate character of agentive awareness.

However, it need not be the case, in order for agentive awareness to subjectively *seem* unmediated, that it actually *is* unmediated. All that is required is that one is not

aware of any inference or observation operative in the source of the thought to the effect that one is A-ing. And if, as I am proposing, drawing the agentive inference is often automatic, it is very likely, just as in the case of many automatic action, that it is rarely if ever a *conscious* inference. In addition, I have argued that agentive awareness is not based on any sensory observation whatsoever, whether conscious or not. So this rules out one of the two ways in which a thought might subjectively seem to one to be mediated. The inferential account predicts that agentive awareness will be subjectively unmediated in the way required.

§5.4 Representational Character of Agentive Awareness

I remarked in Chapter 1 that though there is a consensus as to agentive awareness having representational character, there is far less agreement as to whether this character is on the *thick* or *thin* end of the spectrum (see Bayne & Pacherie, 2007; Pacherie, forthcoming). *Thin* representational content, in this context, merely represents the agent as *acting*, without any further specification of what type of action is being performed. *Thick* representational content, on the other hand, describes further properties of the action, such as the purpose for which it is performed, the way in which it is being performed, and so on. The present account nicely accommodates this lack of consensus about whether the representational character of agentive awareness is thick or thin—for it predicts that agentive awareness is as thick or as thin as the executive intention that it is based upon, and the thickness or thinness of *this* content may vary.

§5.5 Disorders of Volition

How does the present theory make sense of the most common pathologies of agency as described in the clinical literature? I start with anosognosia for hemiplegia

(AHP). Recall that this condition is marked by the inability to acknowledge one's paralysis. Individuals with this condition report performing actions with the paralyzed parts of their bodies, seemingly unaware that they are unable to move in those areas. The inferential account offers a ready explanation for why this occurs. There is evidence that, despite being unable to move the paralyzed parts of their bodies, these individuals are nonetheless able to form executive intentions (see Fotopoulou et al., 2008; Jenkinson & Fotopoulou, 2010). So when such individuals are asked to perform an action with their paralyzed limb, because they do not believe that they are unable to, they may form the executive intention to do so. And because they are not sensitive to their paralysis, they may infer, in the typical way, that they are A-ing on the basis of their executive intention to A. Now in typical cases in which one observes that one is not doing what one forms an executive intention to do, one will abandon the thought that one is A-ing and instead acquire the thought that one has not successfully performed the action one intends to perform. But in the case of individuals with anosognosia for hemiplegia, they do not update the thought that they are A-ing in this way on the basis of this type of observation. Their thought that they are A-ing persists despite countervailing evidence. This interpretation fits comfortably with the inferential account of agentive awareness.

Some in the clinical literature already endorse such an interpretation of AHP (e.g., Vuilleumier, 2004; Davies et al., 2005). And a recent study by Vocat, Saj, Vuilleumier (2013) titled, "The riddle of anosognosia: Does unawareness of hemiplegia involve a failure to update beliefs?" is the first to offer some direct empirical support of this hypothesis. Vocat et al. (2013) had 9 AHP patients as well as 11 healthy controls perform a task in which they were required to guess ten target words (e.g., 'cow'). Each

target word was accompanied by five verbal clues ("I am sometimes black and white," "I am an animal of the female gender"), which gave hints about the semantic or syntactic features of the word. These clues were presented one after the other, with each successive clue providing more information than the previous one. While the first clue left many possible answers, the last clue was meant to leave no doubt as to the identity of the target word. After each clue, the participant would make a guess as to what the target word was, and rate their level of confidence in their guess—they were given no feedback as to the accuracy of their guess. The identity of the target was revealed after each participant's fifth and final guess and confidence rating.

The results were that AHP patients, compared with controls, showed abnormally high confidence in their guesses after the first three clues. Indeed, their confidence ratings after the *first* guess (M = 5.7 on a scale from 0 – 8) were comparable with the ratings that controls gave after their *fourth* guess (M = 5.6). Perhaps even more striking, AHP patients failed to revise their previous guesses even when new information was presented that conflicted with them. The experimenters note that, "[t]hey typically preferred to find 'non-obvious' but 'plausible' connections between the false word provided on a preceding trial rather than reject their current beliefs and make a new guess" (p. 7). In other words, AHP patients would often confabulate reasons for their responses in the same way that they do when it comes to reporting on the actions that they claim to be performing. This study, therefore, offers compelling evidence for the account that the inferential theory of agentive awareness gives of AHP. Anosognosic patients fail to revise or update their thoughts that they are acting in light of new evidence that they are not—instead these thoughts persevere.

What does the inferential account have to say about anarchic hand syndrome? Here, too, the account predicts that there would be no agentive awareness for anarchic hand patients, as there is no executive intention formed to perform the relevant actions. The actions of the anarchic hand patient are at the level of sensorimotor control, not intentional control. They are executed by way of motoric states and sensory feedback, not executive intentions and perceptions. Indeed, as I mentioned, there is evidence that anarchic hand patients are often not aware of the actions of their anarchic hand *until* they observe them. They lack "awareness from the inside", which is provided by thoughts based on executive intentions.

When it comes to explaining utilization behavior, the present account has less to offer. The perplexing thing about these cases is that, unlike in cases of anarchic hand syndrome, these individuals do not seem distressed by their repetitive, stimulus-driven behavior. Indeed, they often confabulate reasons for their behavior insisting, for example, that the experimenter told them to perform the repetitive acts in question. In this way, they exhibit similarities to some cases of anosognosia for hemiplegia. Still, the inferential account might predict that, like those with anosognosia with hemiplegia, patients with utilization behavior have difficulty updating their beliefs about their own agency. More research would need to be carried out in order to garner evidence for or against this proposal.

Finally, I turn to schizophrenic individuals with delusions of control. Within the theoretical framework of the inferential account, there are at least two possible explanations available here: (i) schizophrenic individuals are not aware of their executive intentions to act, and this prevents the downstream inference on the basis of this

awareness from taking place, or (ii) they are unable to draw the inference for some other reason.

Actually, in his earlier writing, Frith (1987) himself proposed something in the spirit of the first explanation, speculating that schizophrenic individuals are selectively impaired in their higher-order awareness of their intentions, or what Frith and others refer to as "self-monitoring." But Frith (2005) later goes on to reject this proposal. It will be useful to quote him at length, in order to get a clear grip on his reasons for doing so:

> In the first formulation of the self-monitoring deficit, I assumed that the patients are not aware of the intended state of their motor system. As a result, their actions, though goal-directed, are a surprise. Consideration of the phenomenology of schizophrenia shows that this idea is wrong. For example, the behaviour of patients with delusions of control is in marked contrast to patients with an anarchic hand (Marchetti & Della Salla, 1998). The Anarchic Hand is a neurological disorder […] that sometimes occurs after lesions to the supplementary motor area (SMA) or the anterior corpus callosum. The contra-lesional hand makes movements, not intended by the patient, like grabbing doorknobs or scribbling with a pencil. The patient is aware that these movements are unintended and tries to stop them. However, such patients are more likely to complain that there is something wrong with their hand, rather than to say that alien forces are controlling their hand. These patients have a problem with the control of one of their hands, but their awareness of what is happening seems to be normal.
>
> Patients with delusions of control do not behave in this way. This is revealed clearly in a study by Sean Spence in which such patients were asked to move a joystick in four different directions at random (Spence et al., 1997). The patients were able to perform this task normally. They made the movements at the correct time and the sequences of movements were as random as those of the controls. And yet the patients reported that their movements were being controlled by outside forces. These patients were making the movements they intended to make since they were successfully following the instructions of the experimenter.

> Furthermore they knew that the movements were intended since they did not try to stop the movements or correct them. Patients with delusions of control do not behave as if their actions were unintended. So why do they experience their movements as being under the control of alien forces and why cannot they make central error corrections? (p. 756)

Frith's worries here may be dealt with by appealing to the distinction between distal intentions and executive intentions. Arguably, what is happening in the case of schizophrenic individuals is that they are aware of their distal intention to do something, say, perform some task according to the experimenter's instructions, but not their executive intention, which is responsible for initiating and guiding that action to completion. For this reason, they report feeling as though their actions are being "pulled out" of them, despite being intended, since they are not aware of any decision to initiate them provided by the executive intention, and they thereby do not come to infer that they are A-ing on this basis.

As for the attributions of agency to external agents, this is a more vexed question that takes us into unknown territory in cognitive psychology concerning the nature of delusions. Again, more research must be done before the present account, or any account for that matter, is able to explain this aspect of schizophrenia.

In this section, I have evaluated the inferential theory of agentive awareness on the basis of the five desiderata that I urged any such theory must meet in Chapter 1. Though the details of the theory remain to be worked out in future research, I hope that its success in meeting these desiderata secures its place as a promising account among the alternatives.

§6 Conclusion

In this chapter, I have motivated, sketched, and evaluated an inferential account of agentive awareness on which one is agentively aware by way of a thought that one is A-ing, which is itself inferred on the basis of an executive intention. I hope to have shown how such a model can account for the self-referential character, the subjectively unmediated character, the phenomenal character, and the representational character of agentive awareness, as well as offer some insight as to what is occurring in prominently discussed disorders of volition. In addition, this model respects the constraints on agentive awareness that I have argued for in Chapters 2, 3, and 4. The further details of the model remain to be fleshed out in future work, but I have laid the groundwork here.

REFERENCES

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: enhancing experienced agency by priming effect information. *Consciousness and Cognition, 14*(3), 439-458.

Anscombe, G. E. M. (1966). *Intention* (2nd ed.). Cambridge, MA: Harvard University Press.

Armstrong, D. M. (1968). *A Materialist Theory of Mind*. London, UK: Routledge & Kegan Paul Ltd.

Assal, F., Schwartz, S., & Vuilleumier, P. (2007). Moving with or without will: functional neural correlates of alien hand syndrome. *Annals of Neurology, 62*(3), 301-306.

Austin, J. L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.

Bach, K. (1975). Performatives are statements too. *Philosophical Studies, 28*(4), 229 - 236.

Baier, A. (1970). Act and Intent. *Journal of Philosophy, 67*(19), 648 - 658.

Banks, G., Short, P., Martinez, J., Latchaw, R., Ratcliff, G., & Boller, F. (1989). The alien hand syndrome: Clinical and postmortem findings. *Archives of Neurology, 46*, 456 - 459.

Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will. In W. P. B. S. Pockett, S. Gallagher (Ed.), *Does consciousness cause behavior?* (pp. 169 - 185). Cambridge, MA: MIT Press.

Bayne, T. (2011a). Agentive experiences as pushmi-pullyu representations. In A. A. B. J. H. Aguilar, K. Frankish (Ed.), *New waves in philosophy of action* (pp. 219 - 236). New York, NY: Palgrave Macmillan.

Bayne, T. (2011b). The sense of agency. In F. Macpherson (Ed.), *The senses* (pp. 490 - 524). Oxford: Oxford University Press.

Bayne, T., & Levy, N. (2006). The feeling of doing: Deconstructing the phenomenology of agency. In W. P. N. Sebanz (Ed.), *Disorders of volition (pp. 49 - 68).* Cambridge, MA: MIT Press.

Bayne, T., & Pacherie, E. (2007). Narrators and comparators: the architecture of agentive self-awareness. *Synthese, 159*, 475 - 491.

Bell, C. (1974). Idea of a new anatomy of the brain. In P. Cranefield (Ed.), *Francois Magendie, Charles Bell and the Course of the Spinal Nerves*. Mt. Kisco, NY: Futura. (Original work published in 1868.)

Bermúdez, J. (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Bermúdez, J. (2010). Action and awareness of agency: Comments on Christopher Frith. *Pragmatics and Cognition, 18*(3), 584 - 596.

Biran, I., Giovannetti, T., Buxbaum, L., & Chatterjee, A. (2006). The alien hand syndrome: What makes the alien hand alien? *Cognitive Neuropsychology, 23*(4), 563-582.

Blakemore, S.-J. (2002). Touching on the cause of delusions. *The Psychologist, 16*(9), 470 - 473.
Blakemore, S.-J., & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology, 13*, 219 - 224.
Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience, 11*(5), 551-559.
Blakemore, S.-J., Goodbody, S. J., & Wolpert, D. M. (1998). Predicting the consequences of our own actions: The role of sensorimotor context estimation. *The Journal of Neuroscience, 18*(18), 7511 - 7518.
Blakemore, S.-J., Oakley, D., & Frith, C. (2001). The neural correlates of volitional awareness. *Emotion and Cognition, 13*(6), S383.
Blakemore, S.-J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol Med, 30*(5), 1131-1139.
Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport, 11*(11), R11-16.
Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1*(7), 635-640.
Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1999). The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *NeuroImage, 10*(4), 448-459.
Blakemore, S. J. (2003). Deluding the motor system. *Consciousness and Cognition, 12*(4), 647-655.
Block, N. (2003). Mental paint. In M. Hahn & B. Ramberg (Eds.), *Reflections and replies: Essays on the philosophy of Tyler Burge*. Cambridge, MA: MIT Press.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci, 30*(5-6), 481-499; discussion 499-548.
Bogen, J. E. (1985). The callosal syndrome. In K. M. Heilman & E. Valentina (Eds.), *Clinical neuropsychology* (2nd ed., pp. 295 - 338). Oxford, UK: Oxford University Press.
Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756.
Brand, M. (1984). *Intending and acting*. Camrbidge, MA: MIT Press.
Bratman, M. (1984). Two Faces of Intention. *The Philosophical Review, 93*(3), 375 - 405.
Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Cambridge University Press.
Bridgeman, B. (2007). Efference copy and its limitations. *Computers in Biology and Medicine, 2007*(37), 924 - 929.
Buehner, M. J. (2012). Understanding the past, predicting the future: causation, not intentional action, is the root of temporal binding. *Psychological Sciences, 23*(12), 1490-1497.
Byrne, A. (2009). Experience and content. *Philosphical Quarterly, 59*(236), 429 - 451.
Carruthers, G. (2012). The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition, 21*, 30 - 45.

Carter, J. D., Bizzell, J., Kim, C., Bellion, C., Carpenter, K. L., Dichter, G., & Belger, A. (2010). Attention deficits in schizophrenia--preliminary evidence of dissociable transient and sustained deficits. *Schizophrenia Research, 122*(1-3), 104-112.

Castañeda, H.-N. (1972). Intentions and intending. *American Philosophical Quarterly, 9*(2), 139 - 149.

Castiello, U., Paulignan, Y., & Jeannerod, M. (1991). Temporal dissociation of motor responses and subjective awareness. A study in normal subjects. *Brain, 114 (Pt 6)*, 2639-2655.

Chalmers, D. (2004). The Representational Character of Experience. In B. Leiter (Ed.), *The Future for Philosophy* (pp. 153-181). Oxford, UK: Oxford University Press.

Chambon, V., Wenke, D., Fleming, S. M., Prinz, W., & Haggard, P. (2013). An online neural substrate for sense of agency. *Cerebral Cortex, 23*(5), 1031-1037.

Chassagnon, S., Minotti, L., Kremer, S., Hoffmann, D., & Kahane, P. (2008). Somatosensory, motor, and reaching/grasping responses to direct electrical stimulation of the human cingulate motor areas. *Journal of Neurosurgery, 109*(4), 593-604.

Chen, Y. C., Thaler, D., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. II. The timing and selection of learned movements. *Exp Brain Res, 102*(3), 461-473.

Cole, J. (1995). *Pride and a daily marathon*. Cambridge, MA: MIT Press.

Cole, J., Gallagher, S., & McNeill, D. (2002). Gesture following deafferentation: A phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences, 1*, 49 - 67.

Danto, A. (1965). Basic actions. *American Philosophical Quarterly, 2*(2), 141 - 148.

Danto, A., & Morgenbesser, S. (1963). What we can do. *The Journal of Philosophy, 60*(15), 435 - 445.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., & Jeannerod, M. (1997). Looking for the agent: An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition, 65*, 71 - 86.

Davidson, D. (2001a). Actions, Reasons, and Causes. *Essays on Actions and Events* (pp. 43 - 61). Oxford, UK: Oxford University Press. (Original published in 1963.)

Davidson, D. (2001b). Agency. *Essays on Actions and Events* (pp. 43 - 61). Oxford, UK: Oxford University Press. (Original published in 1971.)

Davidson, D. (2001c). Freedom to act. *Essays on Actions and Events* (pp. 43 - 61). Oxford, UK: Oxford University Press. (Original published in 1973.)

Davidson, D. (2001d). Intending. *Essays on Actions and Events* (pp. 83 - 102). Oxford, UK: Oxford University Press. (Original published in 1978.)

de Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition, 20*, 82 - 93.

de Vignemont, F. (2012). Bodily immunity to error. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays*. Cambridge, UK: Cambridge University Press.

de Vignemont, F. (forthcoming). A multimodal conception of bodily experiences. *Mind*.

de Vignemont, F., & Fourneret, P. (2004). The sense of agency: A philosophical and empirical review of the ‘‘Who’’ system. *Consciousness and Cognition, 13*, 1 - 19.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227.
Deiber, M. P., Honda, M., Ibanez, V., Sadato, N., & Hallett, M. (1999). Mesial motor areas in self-initiated versus externally triggered movements examined with fMRI: effect of movement type and rate. *Journal of Neurophysiology, 81*(6), 3065-3077.
Della Sala, S. (2005). The anarchic hand. *The Psychologist, 18*(10), 606 - 609.
Dennett, D. C. (1969). *Content and Consciousness*. London, UK: Routledge and Kegan Paul.
Desantis, A., Weiss, C., Schutz-Bosbach, S., & Waszak, F. (2012). Believing and perceiving: authorship belief modulates sensory attenuation. *PLoS One, 7*(5), e37959.
Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4*(11), 423 - 431.
Devitt, M. (2010). Linguistic intuitions revisited. *British Journal of Philosophy of Science, 61*, 833 - 865.
Doniger, G. M., Silipo, G., Rabinowicz, E. F., Snodgrass, J. G., & Javitt, D. C. (2001). Impaired sensory processing as a basis for object-recognition deficits in schizophrenia. *American Journal of Psychiatry, 158*(11), 1818-1826.
Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
Dreyfus, H. (2005). Overcoming the myth of the mental: How philosophers can profit from the phenomenology of everyday expertise. *Proceedings and Addresses of the American Philosophical Association, 79*(2), 47 - 65.
Dreyfus, H. (2007a). Response to McDowell. *Inquiry, 50*(4), 371 - 377.
Dreyfus, H. (2007b). The return of the myth of the mental. *Inquiry, 50*(4), 352 - 365.
Eilan, N., & Roessler, J. (2003). Agency and self-Awareness: Mechanisms and epistemology. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 1 - 47). Oxford, UK: Oxford University Press.
Evans, G. (1982). *The varieties of reference*. New York, NY: Oxford University Press.
Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioral Neurology, 19*(1-2), 53-57.
Farrer, C., Franck, N., Frith, C. D., Decety, J., Georgieff, N., d'Amato, T., & Jeannerod, M. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research: Neuroimaging, 131*, 31 - 44.
Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *NeuroImage, 18*, 324 - 333.
Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. *Consciousness and Cognition, 12*(4), 609-619.
Feinberg, J. (1970). Action and responsibility. *Doing and deserving* (pp. 119 - 151). Princeton, NJ: Princeton University Press.
Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology Review, 15*(2), 73-95.

Flanagan, J. R., & Wing, A. M. (1997). The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. *Journal of Neuroscience, 17*(4), 1519-1528.
Fotopoulou, A., Tsakiris, M., Haggard, P., Vagopoulou, A., Rudd, A., & Kopelman, M. (2008). The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. *Brain, 131*(Pt 12), 3432-3442.
Fourneret, P., Franck, N., Slachevsky, A., & Jeannerod, M. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport, 12*(6), 1203-1208.
Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia, 36*(11), 1133-1140.
Fourneret, P., Paillard, J., Lamarre, Y., Cole, J., & Jeannerod, M. (2002). Lack of conscious recognition of one's own actions in a haptically deafferented patient. *Neuroreport, 13*(4), 541-547.
Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry, 158*, 454 - 459.
Frankfurt, H. G. (1978). The problem of action. *American Philosophical Quarterly, 15*(2), 157 - 162.
Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience, 11*(11), 3656-3666.
Frith, C. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception
and initiation of action. *Psychological Medicine, 17*, 631 - 648.
Frith, C. (1992). *The cognitive neuropsychology of schizophrenia*. East Sussex: Lawrence Erlbaum Associates Ltd.
Frith, C. (2005). The self in action: lessons from delusions of control. *Consciousness and Cognition, 14*(4), 752-770.
Frith, C. (2007). *Making up the mind: How the brain creates our mental world*. Oxford, UK: Blackwell Publishing.
Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition, 21*(1), 52 - 54.
Frith, C., & Done, J. (1989). Positive symptoms of schizophrenia. *British Journal of Psychiatry, 154*, 569-570.
Frith, C. D., Blakemore, S., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research: Brain Research Review, 31*(2-3), 357-363.
Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences, 4*(1), 14 - 21.
Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass, 2*, 347 - 357.
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119 (Pt 2)*, 593-609.
Ginet, C. (1990). *On action*. Cambridge, MA: Cambridge University Press.

Giovannetti, T., Buxbaum, L. J., Biran, I., & Chatterjee, A. (2005). Reduced endogenous control in alien hand syndrome: evidence from naturalistic action. *Neuropsychologia, 43*(1), 75-88.
Gold, J. M., Fuller, R. L., Robinson, B. M., Braun, E. L., & Luck, S. J. (2007). Impaired top-down control of visual search in schizophrenia. *Schizophrenia Research, 94*(1-3), 148-155.
Grice, H. P. (1972). Intention and Uncertainty. Oxford, UK: Oxford University Press.
Guillaud, E., Simoneau, M., & Blouin, J. (2011). Prediction of the body rotation-induced torques on the arm during reaching movements: evidence from a proprioceptively deafferented subject. *Neuropsychologia, 49*(7), 2055-2059.
Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Science, 9*(6), 290-295.
Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience, 9*(12), 934-946.
Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382-385.
Haggard, P., & Cole, J. (2007). Intention, attention and the temporal experience of action. *Consciousness and Cognition, 16*(2), 211-220.
Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research, 126*(1), 128-133.
Haggard, P., & Johnson, H. (2003). Experiences of voluntary action. *Journal of Consciousness Studies, 10*(9 - 10), 72 - 84.
Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *Br J Psychol, 90 (Pt 2)*, 291-303.
Harman, G. (1976). Practical reasoning. *The Review of Metaphysics, 29*(3), 431 - 463.
Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives, 4*, 31-52.
Hécaen, H., Talairach, J., David, M., & Dell, M. B. (1949). Coagulations limitées du thalamus dans les algies du syndrome thalamique: résultats thérapeutiques et physiologiques. *Revue de Neurologies, 81*, 917 - 931.
Hesse, M. D., Nishitani, N., Fink, G. R., Jousmaki, V., & Hari, R. (2010). Attenuation of somatosensory responses to self-produced tactile stimulation. *Cereb Cortex, 20*(2), 425-432.
Hohwy, J., & Frith, C. (2004). Can neuroscience explain consciousness? *Journal of Consciousness Studies, 11*(7-8), 180-198.
Horgan, T. (2012). From agentive phenomenology to cognitive phenomenology: A guide for the perplexed. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 57 - 78). New York: Oxford University Press.
Horgan, T., Tienson, J., George, G. (2003). The phenomenology of first-person agency. In S. Walter & H. Heinz-Dieter (Eds.), *Physicalism and mental causation* (pp. 323 - 340): Imprint Academic.
Humberstone, I. L. (1992). Direction of fit. *Mind, 101*(401), 59 - 83.
Hume, D. (2000). *A treatise of human nature*. Oxford, UK: Oxford University Press. (Original work published in 1739.)

Humphreys, G. W., & Riddoch, M. J. (2003). Fractionating the intentional control of behaviour: A neuropsychological analysis. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 201 - 217). Oxford, UK: Oxford University Press.

James, W. (2007). *Principles of Psychology, Vol 1*. New York, NY: Cosimo Inc. (Original work published in 1890.)

Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford, UK: Blackwell Publishers, Inc.

Jeannerod, M. (2006). *Motor Cognition: What actions tell the self*. New York, NY: Oxford University Press.

Jenkins, I. H., Jahanshahi, M., Jueptner, M., Passingham, R. E., & Brooks, D. J. (2000). Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. *Brain, 123 (Pt 6)*, 1216-1228.

Jenkinson, P. M., & Fotopoulou, A. (2010). Motor awareness in anosognosia for hemiplegia: experiments at last! *Experimental Brain Research, 204*(3), 295-304.

Keeley, B. L. (2002). Making sense of the senses: Individuating modalities in humans and other animals. *Journal of Philosophy, 99*(1), 5-28.

Knobe, J. (2003). Intentional action and side-effects in ordinary language. *Analysis, 63*, 190 - 193.

Knoblich, G., & Repp, B. H. (2009). Inferring agency from sound. *Cognition, 111*(2), 248-262.

Lafargue, G., & Franck, N. (2009). Effort awareness and sense of volition in schizophrenia. *Consciousness and Cognition, 18*(1), 277-289.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science, 303*, 1208 - 1210.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. *Journal of Neuroscience, 26*(27), 7265-7271.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2007). Manipulating the experienced onset of intention after action execution. *Journal of Cognitive Neuroscience, 19*(1), 81-90.

Leudar, I., Thomas, P., & Johnston, M. (1994). Self-monitoring in speech production: effects of verbal hallucinations and negative symptoms. *Psychological Medicine, 24*(3), 749-761.

Levy, N., & Bayne, T. (2004). A will of one's own: Consciousness, control, and character. *International Journal of Law and Psychiatry, 27*, 459 - 470.

Lhermitte, F. (1983). 'Utilization behaviour' and its relation to lesions of the frontal lobes. *Brain, 106 (Pt 2)*, 237-255.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain, 106 (Pt 3)*, 623-642.

Lindner, A., Thier, P., Kircher, T. T., Haarmeier, T., & Leube, D. T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology, 15*(12), 1119-1124.

Locke, J. (1975). *An essay concerning human understanding*. Oxford, UK: Clarendon Press. (Original published in 1689.)

Logan, G. D., & Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science, 330*(6004), 683-686.
Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
MacPherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research, 84*(1), 24 - 62.
Malaspina, D., Simon, N., Goetz, R. R., Corcoran, C., Coleman, E., Printz, D., Mujica-Parodi, L., & Wolitzky, R. (2004). The reliability and clinical correlates of figure-ground perception in schizophrenia. *Journal of Neuropsychiatry and Clinical Neuroscience, 16*(3), 277-283.
Mandik, P. (2010). Control consciousness. *Topics in Cognitive Science, 2*, 643 - 657.
Marcel, A. (2003). The sense of agency: Awareness and ownership of action. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 48 - 93). Oxford: Oxford University Press.
Marchetti, C., & Della Sala, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry, 3*(3), 191 - 207.
Marteniuk, R. G., MacKenzie, C. L., Jeannerod, M., Athenes, S., & Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology, 41*(3), 365-378.
Mather, G., Pavan, A., Campana, G., & Casco, C. (2008). The motion aftereffect reloaded. *Trends in Cognitive Science, 12*(12), 481-487.
McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.
McDowell, J. (2007). What myth? *Inquiry, 50*(4), 338 - 351.
McDowell, J. (2010). What is the content of an intention in action? *Ratio, 23*, 415 - 432.
Mele, A. (1992). *Springs of action*. New York: Oxford University Press.
Mele, A., & Moser, P. K. (1994). Intentional action. *Noûs, 28*(1), 39 - 68.
Mellors, C. S. (1970). First-rank symptoms of schizophrenia. *British Journal of Psychiatry, 117*, 15 - 23.
Metzinger, T. (2006). Conscious volition and mental Representation: Toward a more fine-grained analysis. In N. Sebanz & W. Prinz (Eds.), *Disorders of volition* (pp. 19 - 49). Cambridge, MA: MIT Press.
Michotte, A. (1946). *La perception de la causalité*. University of California: Editions de l'Institut Supérieur de Philosophie.
Montero, B. (2010). Does bodily awareness interfere with highly skilled movement? *Inquiry, 53*(2), 105 - 122.
Montero, B., & Evans, C. A. (2011). Intuitions without concepts lose the game: mindedness in the art of chess. *Phenomenology and Cognitive Science, 10*, 175 - 194.
Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition, 18*(4), 1056-1064.
Mylopoulos, M. I. (2011). Why reject a sensory imagery theory of control consciousness? *Topics in Cognitive Science, 3*(2), 268 - 272.
Mylopoulos, M. I., & Lau, H. (forthcoming). Naturalizing free will: Paths and pitfalls. In A. Mele (Ed.), *Surrounding free will*. New York, NY: Oxford University Press.

Mylopoulos, M. I., & Pereplyotchik, D. (2013). Is there any evidence for forward modeling in language production? *Behavioral and Brain Sciences*, 36(4), 368 - 369.

Nachev, P., Wydell, H., O'Neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. *NeuroImage, 36*(3-3), T155 - T163.

Nahmias, E. (2005). Agency, authorship, and illusion. *Consciousness and Cognition, 14*, 771 - 785.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004). The phenomenology of free will. *Journal of Consciousness Studies, 11*(7 - 8), 162 - 179.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, R. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1 - 18). New York, NY: Plenum Press.

O'Shaughnessy, B. (1998). Proprioception in the body image. In J. L. Bermúdez, A. J. Marcel, & N. Eilan (Eds.), *The body and the self* (pp. 175 - 204). Cambridge, MA: MIT Press.

Pacherie, E. (2000). The content of intentions. *Mind & Language, 15*(4), 400 - 432.

Pacherie, E. (2006). Toward a dynamic theory of intentions. In W. P. B. S. Pockett, S. Gallagher (Ed.), *Does consciousness cause behavior?* (pp. 145 - 168). Cambridge, MA: MIT Press.

Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche, 13*(1), 1 - 30.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*, 179 - 217.

Pacherie, E. (2011). Nonconceptual representations for action and the limits of intentional control. *Social Psychology, 42*(1), 67 - 73.

Pacherie, E. (forthcoming). Time to act: The dynamics of agentive experiences. *Philosophical Topics*.

Pacherie, E., & Haggard, P. (2010). What are intentions? In L. Nadel & W. Sinnott-Armstrong (Eds.), *Benjamin Libet and agency* (pp. 70 - 84). Oxford, UK: Oxford University Press.

Paul, S. (2009). How We Know What We're Doing. *Philosophers' Imprint, 9*(11), 1 - 24.

Peacocke, C. (2001). Does Perception Have a Nonconceptual Content? *Journal of Philosophy, 98*, 239-264.

Penfield, W. (1975). *The mystery of the mind: A critical study of consciousness and the human brain*. Princeton, NJ: Princeton University Press.

Perner, J. (2003). Dual control and the causal theory of action: The case of non-intentional action. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 218 - 243). Oxford, UK: Oxford University Press.

Perry, J. (1993). The Problem of the Essential Indexical *The problem of the essential indexical and other essays*. New York: Oxford University Press.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral Brain Sciences*, 1-19.

Pitt, D. (2004). The phenomenology of cognition or what is it like to think that p? *Philosophy and Phenomenological Reserach, 69*(1), 1 - 36.

Prinz, J. J. (2007). All consciousness is perceptual. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind*. Malden, MA: Blackwell Publishing Ltd.

Prinz, J. J. (2012). *The conscious brain: How attention engenders experience*. New York, NY: Oxford University Press.

Purkinje, J. (1825). Über die Scheinbewegungen, welche im subjectiven Umfang des Gesichtsinnes vorkommen. *Bulletin der naturwessenschaftlichen Sektion der Schlesischen Gesellschaft, 4*, 9 - 10.

Reason, J. T. (1990). *Human Error*. New York, NY: Cambridge University Press.

Reid, T. (1997). An Inquiry into the Human Mind on the Principles of Common Sense. D. R. Brookes (Ed.). University Park, PA: The Pennsylvania University Press. (Original work published in 1764.)

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research: Cognitive Brain Research, 3*(2), 131-141.

Rosenthal, D. (2012). Awareness and identification of self. In J. Liu & J. Perry (Eds.), *Consciousness and the self: New essays*. Cambridge, UK: Cambridge University Press.

Rosenthal, D. M. (2002). The timing of conscious states. *Consciousness and Cognition, 11*(2), 215-220; discussion 304-217, 326-233.

Rosenthal, D. M. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology* (pp. 17 - 44). Philadelphia, PA: John Benjamins B. V.

Rosenthal, D. M. (2005). *Consciousness and mind*. New York: Oxford University Press.

Rosenthal, D. M. (2010). How to think about mental qualities. *Philosophical Issues, 20*, 368 - 387.

Roskies, A. (2010). Freedom, neural mechanism, and consciousness. In A. R. M. R. F. Baumeister, K. D. Vohs (Ed.), *Free will and consciousness: How Might They Work* (pp. 153 - 171). Oxford, NY: Oxford University Press.

Ryle, G. (2000). *The concept of mind*. Chicago, IL: University of Chicago Press.

Sacks, O. (1998). *The man who mistook his wife for a hat: And other clinical tales*: Touchstone.

Sarlegna, F. R., & Bernier, P. M. (2010). On the link between sensorimotor adaptation and sensory recalibration. *Journal of Neuroscience, 30*(35), 11555-11557.

Sarlegna, F. R., Gauthier, G. M., Bourdin, C., Vercher, J. L., & Blouin, J. (2006). Internally driven control of reaching movements: a study on a proprioceptively deafferented subject. *Brain Research Bulletin, 69*(4), 404-415.

Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, 74 - 83.

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Science U S A, 109*(42), E2904-2913.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge, MA: Cambridge University Press.

Searle, J. R. (1989). How Performatives Work. *Linguistics in Philosophy, 12*(5), 535 - 558.

Sebanz, N., & Prinz, W. (2006). *Disorders of volition*. Cambridge, MA: MIT Press.
Sellars, W. (1976). Volitions Re-affirmed. In D. W. M. Brand (Ed.), *Action theory* (pp. 47 - 66). Hingham, MA: D. Reidel Publishing Company, Inc.
Setiya, K. (2007). *Reasons without rationalism*. Princeton, NJ: Princeton University Press.
Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry, 162*(12), 2384-2386.
Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy, 65*(19), 555 - 567.
Shoemaker, S. (1994). Self-knowledge and "inner sense": Lecture I: The object perception model. *Philosophy and Phenomenological Reserach, 54*(2), 249 - 269.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*(5), 543-545.
Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120 (Pt 11)*, 1997-2011.
Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Computational Physiology and Psychology, 43*(6), 482-489.
Stirling, J. D., Hellewell, J. S. E., & Ndlovu, D. (2001). Self-monitoring dysfunction and the positive symptoms of schizophrenia. *Psychopathology, 34*, 198 - 202.
Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain, 120 (Pt 3)*, 535-559.
Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain, 133*(Pt 1), 262-271.
Synofzik, M., Vosgerau, G., & Lindner, A. (2009). Me or not me - An optimal intergration of agency cues? *Consciousness and Cognition, 18*, 1065 - 1068.
Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219 - 239.
Thaler, D., Chen, Y. C., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Experimental Brain Research, 102*(3), 445-460.
Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology, 77*, 533-546.
Trillenberg, P., Heide, W., Junghanns, K., Blankenburg, M., Arolt, V., & Kompf, D. (1998). Target anticipation and impairment of smooth pursuit eye movements in schizophrenia. *Experimental Brain Research, 120*(3), 316-324.
Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research, 149*(4), 439-446.
Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition, 96*(3), 215-231.
Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.

Tye, M. (2003). *Consciousness and persons: Unity and identity*. Cambridge, MA: MIT Press.
Velleman, J. D. (1989). *Practical reflection*. Princeton, NJ: Princeton University Press.
Vendler, Z. (1972). *Res cogitans: An essay in rational psychology*. Ithaca, NY: Cornell University Press.
Vocat, R., Saj, A., & Vuilleumier, P. (2013). The riddle of anosognosia: Does unawareness of hemiplegia involve a failure to update beliefs? *Cortex, 49*(7), 1771-1781.
Von Helmholtz, H. (1866). *Handbuch der physiologischen Optik*. Leipzig: Voss.
Von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. Wechselwirkungen zwischen Zerntalnervensystem und Peripherie. *Naturwissenschaften, 27*, 464 - 476.
Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience, 9*(1), 26-27.
Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain, 133*(10), 3104-3112.
Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: Bradford Books.
Wegner, D. (2005). Who is the controller of controlled processes? In J. S. U. R. R. Hassin, J. A. Bargh (Ed.), *The new unconscious* (pp. 19 - 36). New York, NY: Oxford University Press.
Wegner, D. M. (2004). Precis of the illusion of conscious will. *Behavioral and Brain Sciences, 27*(5), 649-659; discussion 659-692.
Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86*(6), 838-848.
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *American Psychologist, 54*(7), 480-492.
Weiller, C., Juptner, M., Fellows, S., Rijntjes, M., Leonhardt, G., Kiebel, S., Muller, S., Diener, H. C., & Thilmann, A. F. (1996). Brain representation of active and passive movements. *Neuroimage, 4*(2), 105-110.
Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary observations on tickling oneself. *Nature, 230*(5296), 598-599.
Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition, 115*(1), 26-38.
Wing, A. M., Flanagan, J. R., & Richardson, J. (1997). Anticipatory postural adjustments in stance and grip. *Experimental Brain Research, 116*(1), 122-130.
Wittgenstein, L. (2001). *Philosophical Investigations* (3rd ed.). Malden, MA: Blackwell Publishers Ltd. (Original work published in 1953.)
Wohlschläger, A., Engbert, K., & Haggard, P. (2003). Intentionality as a constituting condition for the own self—and other selves. *Consciousness and Cognition, 12*, 708 - 716.
Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cogn Science, 1*(6), 209-216.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880-1882.
Wolpert, D. M., & Miall, R. C. (1996). Forward models for physiological motor control. *Neural Networks, 9*(8), 1265-1279.
Wong, H. Y. (2012). A measure of my agency? *Consciousness and Cognition, 21*(1), 48-51.