APPROVAL

Name: Myrto Ioanna Mylopoulos
Degree: Master of Arts in the Department of Philosophy

Title of Thesis: The Experience of Agency and the Self

Examining Committee:

Chair: Dr. Martin Hahn
Associate Professor, Department of Philosophy

Dr. Kirstie Laird
Senior Supervisor
Assistant Professor, Department of Philosophy

Dr. David Zimmerman
Supervisor
Professor, Department of Philosophy

Dr. Mark Blair
External Examiner
Assistant Professor, Department of Psychology

Date Defended/Approved: August 11, 2006
DECLARATION OF
PARTIAL COPYRIGHT LICENCE

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection, and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
ABSTRACT

The main focus of this thesis is the conscious experience that attends the actions that we perform as agents. The ‘feeling of doing’ is widely acknowledged to be an essential component of our mental life, yet there has been little effort to study it seriously up until fairly recently. The first half of the thesis is devoted to a discussion of some metaphysical issues surrounding the veridicality of the experience of agency. I start by evaluating the arguments of ‘agent-causationists’ who, within the more general framework of their libertarian theory of free will, purport to demonstrate that the experience of agency is veridical and can be directly translated into a naturalistically plausible metaphysics. Next, I turn to the ‘illusionists’ who attempt to show that, in one way or another, the experience of agency is misleading. I argue that both camps fail to make their case effectively. The second half of this thesis is concerned with the phenomenology of the experience of agency. I sketch an account of its content and go on to argue that, described as such, it may be viewed as a rich source of self-knowledge that plays an indispensable role in the creation and development of our self-concept.
DEDICATION

To the philo.
ACKNOWLEDGEMENTS

Thank you first and foremost to my senior supervisor/friend Kirstie Laird for her two years of support, encouragement, kindness and wisdom. Thanks also to Dr. Z and to Mark Blair for being a part of this process.

I am also indebted to my fellow grads for keeping me somewhat sane during all of this. I’d like to especially thank Lee-Anna Sangster for always being there for me and for thinking rationally when I could not and Jenn Warriner for listening.

I have to also thank Louise Norman for all her help in making sure things got done when they needed to be done.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approval</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Agent-causationists: The experience of agency as veridical</td>
<td>5</td>
</tr>
<tr>
<td>Wegner and Libet: The experience of agency as illusion</td>
<td>24</td>
</tr>
<tr>
<td>A sketch of the phenomenology</td>
<td>46</td>
</tr>
<tr>
<td>The experience of agency and the self</td>
<td>61</td>
</tr>
<tr>
<td>Conclusion</td>
<td>72</td>
</tr>
<tr>
<td>Reference list</td>
<td>75</td>
</tr>
</tbody>
</table>
INTRODUCTION

There is a distinct phenomenal difference between the experience of performing an intentional act and the experience of making an unintentional movement. The aim of my thesis is to take this qualitative difference seriously and examine some important issues surrounding it. Far from collapsing this particular problem into that of understanding conscious experience in general, I wish to treat 'the experience of agency' as an area of inquiry in its own right.

Indeed, the conviction that the experience of agency is a phenomenon that warrants individual attention is widely held, as evidenced by statements touting it as “an important dimension of mental life that has been largely overlooked in recent philosophy of mind” (Graham, Tienson & Horgan, 2003); “essential to our general understanding of free will, moral responsibility and human agency” (Zhu, 2003b); and “fundamental to our sense of self” (Haggard et al., 2002). Despite these endorsements, significant mention of the subjective experience of acting is conspicuously absent, most notably so in an area of philosophical research in which one would expect it to play a dominant role: action theory. It will be useful to begin by offering a cursory glance at a prominent account of agency, viz. that offered by Donald Davidson, both to get clearer on the concept, and to highlight the importance of questions concerning its subjective component.

What is an action? This question turns out to be harder to answer than one might expect, for we must be able to distinguish between bodily movements (e.g. kicking a ball) that we willingly perform from bodily movements that we do not willingly perform (e.g.
a knee-jerk reflex). One candidate for a criterion with which to determine whether or not a particular bodily movement is an action is being intentional. With this in mind, we can easily see why, in most cases, for example a knee-jerk would not count as an action and kicking a ball would.

However, one must be careful here, for in certain cases, it seems possible to say that not all actions are intentional. Consider Davidson’s (1971) example of spilling the contents of one’s cup. If I pour out the contents of my cup, falsely believing that it is filled with tea when in actuality it is filled with coffee, then I have poured out the coffee, which is an action, but not an intentional one. Similarly, when Lois Lane kisses Clark Kent, thinking he is Superman, she does not do so intentionally, but there is no question that she performs an action. On the other hand, as Davidson points out, if, instead of me pouring out the contents of my cup, someone had jiggled my hand and I had then spilled the coffee, nobody would argue that I had performed an action. The relevant difference between me dropping my coffee after my hand is jiggled and me pouring out my coffee seems to lie in the fact that under a particular description, what I have done in the latter case can be described as intentional, even if there are descriptions under which it is not. Thus, pouring out the contents of my cup and Lois Lane kissing the man in front of her are both actions because they are things that are done intentionally under those descriptions, regardless of the fact that we could describe these doings as unintentional by saying that I poured out my coffee and Lois Lane kissed Superman\textsuperscript{1}.

There is now the further difficult question of what makes an action, under a certain description, intentional. According to Davidson (1963), intentional actions are

\textsuperscript{1} As Davidson points out, this commits us to talk of descriptions of actions rather than actions, otherwise we would be able to say that a particular action was both intentional and not intentional.
those that can be explained by appealing to the agent’s reasons for performing them. These reasons are nothing more than pairs of pro-attitudes (e.g. desires) and beliefs, which serve as causes of the agent’s actions. For example, if I go to the store to get milk, this action is intentional in virtue of the fact that I do it for the reason that I want milk, which can be broken down into my desire for milk and my belief that going to the store will be a way of fulfilling this desire.

Whether or not one accepts Davidson’s particular way of characterizing actions, we now have a reasonable answer to the original question we asked: actions are those bodily movements that are intentional under a certain description. But what becomes apparent at this point is the lack of appeal to the agent’s experience of acting as a relevant factor for determining whether or not an action has been done intentionally. It is fairly evident that the experience of agency is not something that can be ignored in any discourse on action, since it seems so natural to view the experience as the most reliable marker of agency and action. After all, one’s movements may seem goal-directed or even rational, but they common sense suggests that they are not actions unless the agent experiences them as such ‘from the inside’. As the psychologist Daniel Wegner nicely puts it,

[i]f a person plans to take a shower, for example, and says that she intends to do it as she climbs into the water, spends 15 minutes in there scrubbing up nicely, and then comes out reporting that she indeed seems to have had a shower—but yet also reports not feeling she had consciously willed her showering—who are we to protest? Consciously willing an action requires a feeling of doing [...], a kind of internal “oomph” that somehow certifies authentically that one has done the action. If the person did not get that feeling about her shower, then even if we climbed in with her to investigate, there is no way we could establish for sure whether she consciously willed her showering (Wegner, 2004, p. 650).
The close intuitive connection between the experience of agency and action, highlighted above, raises several questions in need of answering. These questions, I think, can fall under two categories: those pertaining to the veridicality of the experience of agency and those pertaining to its phenomenological content.

In the first two chapters of this thesis, I will focus on the questions surrounding the veridicality of the experience of agency, i.e. whether or not it accurately represents us as the causes of our actions, by evaluating the arguments of ‘agent-causationists’ and ‘illusionists’, the former wishing to show that the experience of agency is an accurate representation of the way that we actually exercise our agency and the latter wishing to show that it is illusory in some sense. I will argue that neither group succeeds in defending its stance beyond reasonable doubt. In the second half of this thesis I shift gears and examine the phenomenology of the experience of agency. I sketch an account of its phenomenological content, and I conjecture as to how the experience, described as such, might constitute a source of self-knowledge and aid in the creation and development of our self-concept.
AGENT-CAUSATIONISTS: THE EXPERIENCE OF AGENCY AS VERIDICAL

A natural way of characterizing our typical experience of making decisions and acting upon them – one that would, I think, gain widespread assent – goes something like this: When I decide, say, to go for a walk on a cool autumn evening, I am conscious of various factors at work (some consciously articulated, some not) motivating me either to do so or to do something else instead. And there are some courses of action which, while it is conceivable that I might choose to follow them, are such that they do not represent ‘genuine’ possibilities for me at that time, given my current mood, my particular desires and beliefs, and, in some cases, long-standing intentions of a general sort. But within the framework of possibilities (and perhaps even relative likelihoods) that these present conative and cognitive factors set, it seems for all the world to be up to me to decide which particular action I will undertake. The decision I make is no mere vector sum of internal and external forces acting upon me during the process of deliberation (if, indeed. I deliberate at all). Rather, I bring it about – directly, you might say – in response to the various considerations: I am the source of my own activity, not merely in a relative sense as the most proximate and salient locus of an unbroken chain of causal transactions leading up to this event, but fundamentally, in a way not prefigured by what has gone before. Or, again, so it seems (O’Connor, 1995, p.173).

Probably the strongest formal attempt to vindicate and remain faithful to the experience of acting described above, that is, the experience that I am the source and ultimate cause of my actions, comes in the form of theories of agent-causation. These theories are generally defended in the context of the free will problem as viable options for libertarians wishing to escape the clutches of determinism. My primary focus, however, will not be on their philosophical merit as libertarian accounts of free will, although this may indeed depend partly on my evaluation of them here. Rather, I am interested in exploring how proponents of these theories weave the threads of our
experience into a metaphysical fabric that they take to be without holes. In this section, I aim to evaluate the claim that our experience of agency can be translated directly into a naturalistically plausible metaphysics, that is, a metaphysics consistent with everything we know about the workings of the natural world.

Agent-causation theories take as their starting point the intuition that I, and not some event in my brain, am the sole cause of my voluntary actions. In order to accommodate this special intuition, agency theorists make the bold move of positing an entirely novel and primitive form of causation, which obtains between an agent and an event and is distinct from and irreducible to event-causation, which obtains between two events. It is important to note the radical nature of a commitment to the possibility of agent-causation, for it is a commitment to a type of causal relationship that is restricted entirely to human action, and appears nowhere else in nature. An agent, on this view, is a *sui generis* cause, or, as Roderick Chisholm more boldly puts it, “we have a prerogative which some would attribute only to God: each of us when we act, is a prime mover unmoved. In doing what we do, we cause certain things to happen, and nothing – or no one- causes us to cause those events to happen” (Chisholm, 1964, p.152).

Agent-causation derives its plausibility from the fact that it fits so well with our experience of acting. Indeed, in the past this experience has been taken as the *only* evidence for the existence of any kind of causation at all. Bishop Berkeley, for example, puts so much weight on our experiences of causing our actions that he claims that the sole type of causation in the world is agent-causation: “Thus much is certain, and grounded on experience: but when we talk of unthinking agents, or of exciting ideas exclusive of volition, we only amuse ourselves with words” (Berkeley, 1790).
Although agent-causationists are not committed to the ubiquity of the agent cause, there are many who nonetheless challenge the very coherence of agent-causation within a naturalistic framework. Reactions to agent-causation range from outright incredulity, e.g. “from this philosopher’s armchair this idea looks patently crazy” (Flanagan, 2002, p.106), to deep suspicion:

How does an agent cause an effect without there being an event (in the agent, presumably) that is the cause of that effect (and is itself the effect of an earlier cause, and so forth)? Agent causation is a frankly mysterious doctrine, positing something unparalleled by anything we discovered in the causal processes of chemical reaction, nuclear fission and fusion, magnetic attraction, hurricanes, volcanos, or such biological processes as metabolism, growth, immune reactions, and photosynthesis (Dennett, 2003, p.100).

Of course, these skeptical reactions alone do not demonstrate that agent-causation is, in fact, an incoherent notion. In order to determine whether this is or is not the case, we must examine the agent-causationist’s proposal more carefully.

So how is agent-causation supposed to work? Surprisingly enough, the details of the agent as cause are seldom offered, despite the evident fact that a coherent metaphysics is necessary in order to get agent-causation off the ground at all. Proponents of agent-causation are most often concerned with other details, all of which take the intelligibility of an agent-cause for granted. These include what it is that the agent causes - i.e. a determination of the will (Reid, 1969), a neurophysiological event (Chisholm, 1964), an intentional state (O’Connor, 2000a), an action (Clarke, 1996), etc. - the role of reasons in the production of free action, whether or not the event of the agent’s causing an event is itself caused, and so on and so forth. Although these are all important questions that must be answered by any agent-causationist, I take the question of whether
an agent can be a cause in the first place to be prior and of chief importance for the vindication of the experience of agency. In what follows, I will consider various difficulties that arise in attributing causal powers to agents rather than events.

Traditionally, agent-causation is posited as a primitive form of causation that is exercised by purposive agents in the production of free actions, and is distinct from and irreducible to event causation. These theories exclude the possibility that chains of events governed by natural laws, whether indeterministically or deterministically conceived, provide the crucial causal springs behind our free actions; that role is reserved for the agent alone (see Reid, 1788; Chisholm, 1964; Taylor, 1966). In taking event-causation out of the equation, agent-causationists see themselves as shifting the focus of the free will debate back to the agent, and away from events involving the agent, such as reasons or intentions, that have typically received the most attention in attempts to offer compatibilist accounts of the production of free actions. This strategy allows agent-causationists to avoid both horns of a dilemma faced by free will theorists in general.

The crux of the dilemma is that event-causation entails that our free actions are caused either indeterministically or deterministically by antecedent events. If indeterministically, then our moral responsibility rests on little more than an Epicurean swerve, perhaps a random quantum event that causes us to perform action A as opposed to action B, but this hardly seems a robust enough foundation for blame and praise, reward or punishment. On the other hand, if our actions are deterministically caused by prior events, their causal origins can be traced through chains of events back to before we were even born, in which case ascriptions of moral responsibility seem likewise unwarranted. Of course, debate surrounding this dilemma abounds, and numerous
arguments proffering solutions in response to each of its horns have been formulated, some of which may indeed prove successful. My main point here is that the agent-causationist aims to avoid the dilemma entirely by denying that any of our free actions are caused by events, either deterministically or indeterministically.

Although this evasive manoeuvre may prove successful for dealing with the aforementioned dilemma, theories of agent-causation are faced with additional problems of their own. For one thing, they carry the burden of dissolving the mystery that surrounds the notion of an agent, rather than an event, acting as a cause. It turns out that agent-causationists are often tight-lipped when it comes to expanding upon the metaphysical status of the agent. This reticence does not serve them well, for it leaves their accounts open to the objection that they are substance dualist accounts in disguise, positing a different kind of substance altogether as the source of our free actions in a vain attempt to save human freedom and moral responsibility. If they do indeed endorse substance dualism as one of their key tenets, then they are vulnerable to a number of objections that have been launched against this doctrine, one or more of which may prove fatal to their account.²

Regardless of whether or not a substance dualist agency theory can be successfully defended, contemporary agent-causationists are not interested in doing so; rather they are after an empirically viable account of how agents cause actions that does not require an appeal to immaterial entities in order to function coherently. For this

² I mention in passing a few possible objections to any substance dualist account of free will, namely the problems inherent in explaining how a non-spatial, non-temporal entity might affect the movement of physical bodies at particular times, how one might individuate these entities either from the first or third person and moreover how it is that they are non-arbitrarily anchored to one particular body rather than another.
reason, I will proceed on the assumption that an agent is a biological organism - a human being - and that any causal powers it possesses are describable in naturalistic terms.\(^3\) Whether or not this assumption is warranted and the consequences of its denial are issues that may be debated elsewhere.

Even if working on the assumption that an agent is a naturally occurring biological organism, as I am here, the very idea of agent-causation remains rather mysterious, for how might a human being as a whole be a cause of events internal to it? One of the earliest and most frequently cited attempts at answering this question comes from Roderick Chisholm (1964). As mentioned earlier, Chisholm supposes that in the production of actions, the agent acts as some sort of “prime mover unmoved”, an uncaused cause of an event in the brain that sets in motion a chain of events culminating in a free act. He describes the situation as follows: “... whenever a man does something A, then (by ‘immanent causation’)\(^4\) he makes a certain cerebral event happen, and this cerebral event (by ‘transcendent causation’)\(^5\) makes A happen” (Chisholm, 1964, p.149). On Chisholm’s account we are not to view the agent’s causing a cerebral event as a ‘bringing about’, that is, the agent does not do something to cause it, like, for example, one might strike a match to bring about fire. Rather, the agent as a whole stands in a direct causal relation to the cerebral event.

This claim leaves one puzzling over how this ‘direct causal relation’, given its unique nature, might be unpacked. Chisholm himself anticipates the objection that “there

\(^3\) This may not be such a bold assumption. Chisholm (1964) uses the terms ‘agent’ and ‘man’ interchangeably, and Randolph Clarke’s agent is a "living organism, understood to be a physical substance" (Clarke, 1996, p.44 fn1)

\(^4\) Chisholm uses this term to denote the type of causation between an agent and an event (Chisholm, 1964, p.148)

\(^5\) Chisholm uses this term to denote the type of causation between events (Chisholm, 1964, p.148)
is no more to the man’s action in causing event A than there is to the event A’s happening by itself” (Chisholm, 1964, p.150). Chisholm replies that there is indeed a difference between the man’s action causing event A and event A’s happening by itself, and that is simply that, in the former case, event A was caused by the man, and in the latter case it was not (Chisholm, 1964, p.151). Whatever shortcomings this explanation has in the context of agent-causation, Chisholm argues, may also be attributed to an analogous explanation in the putatively unproblematic context of event-causation. After all, according to Chisholm, the only difference between saying of two events A and B, that B happened and then A happened, and that event B caused event A, is that in the latter case, event A had a cause, and that was event B.

However, there are significant differences between agents and events that make an analogy along these lines more problematic than Chisholm suggests, and expose deeper issues that need to be resolved in positing an agent as a cause. For in the case of two events, there are ways to understand the causal relationship between them that cannot possibly be employed in the case of an agent and an event. I will consider one substantive difference between the two that is often mentioned, namely that an event is a datable entity, whereas an agent is not. The following objection, put forward by C.D. Broad, is motivated by this observation:

I see no prima facie objection to there being events that are not completely determined. But, in so far as an event is determined, an essential factor in its total cause must be other events. How can an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event? (as quoted in O’Connor, 2000b, p.74)
Broad's objection points to a metaphysical feature of the agent-causal relationship that is in need of defence; for one thing that agent-causationists have trouble explaining is why a particular event that is the effect of an agent-cause might occur at one particular time rather than another, since it does not make sense to say of an agent, supposedly the sole cause of that event, that it occurs at any time at all.

One way of getting around this objection has been offered by Randolph Clarke (1996), who propounds what he calls the "causal agent-causal view", which is a hybrid view allowing a role for both event-causation and agent-causation in the production of free action. The picture is as follows: An agent is capable of reflecting rationally on her alternative courses of action and is motivated to act by her reasons. Furthermore, when an agent performs a free action, she exercises a certain causal power from within her rational framework, that is distinct from and irreducible to any events involving the agent. Events themselves are construed as nondeterministic causes, leaving open the possibility of alternative effects resulting from the same event-cause. In this way the agent is free to cause any one of the possible actions left open by the nondeterministic event causes, which may include certain of the agent's reasons. Although prior events may determine that the agent will act, they do not determine in what way the agent will act. On this account then, the agent cause is the determining factor in the production of any given free action, while antecedent events are not denied a causal role as they are in traditional agency theories.

By countenancing events as partial causes of our free actions, the causal agent-causal view may appeal to event causes in an explanation of why an agent-caused event occurred at one time rather than another (Clarke, 1996, p.40). Clearly this is an
advantage not held by traditional agency theories. However, although the causal agent-causal view is indeed one way to handle Broad’s objection, the way in which it does so, i.e. by appealing to event-causation, simply draws our attention to the fact that all existing models of causation require that causes be events, and makes agent-causation seem all the more suspicious. Moreover, Clarke’s causal agent-causal theory is not equipped to deal with another threat to the intelligibility of agent-causation that stems from the reluctance of most agency theorists to offer an explanation of the property or set of properties, if any, from which the special causal powers of an agent derive.

In general, on a given model of event-causation, we take an event to cause another event in virtue of the properties possessed by the object(s) involved in each one. We say that the strike of lightning is the cause of the tree’s falling because of the special law-like relation between certain properties of strikes of lightning and certain properties of trees. It seems likely that if an agent is to be the cause of an action, it must be in virtue of a property or a set of properties that the agent possesses. This is especially so since in certain cases we do not consider a human being to be an agent at all, for example in the case of one who is severely mentally ill or an infant. There must be some property that these human beings lack that an agent possesses, or else we would not be warranted in excluding them from the class of things that can be considered agents. Thus, the agent-causationist owes an explanation of the property or the set of properties in virtue of which a human being is able to act as an agent-cause.

That being said, agent-causationists must avoid identifying the agent-causal power in question with any property that may be reducible to a physical property of the brain, for then its activation would be governed by physical laws, and there would be
little room for libertarian free will of the kind sought after by the agency theorist. On the other hand, the agent-causal power must be grounded in the natural world for reasons I have already gone over, putting the agent-causationist in the same tough spot that many non-reductive physicalists, who propose that mental properties are dependent on, yet irreducible to physical properties, find themselves. The agent-causationist, however, is actually worse off than the non-reductive physicalist in that he also requires that this property provide its possessor with an entirely novel form of causal power that does not occur anywhere else in nature. What is the agency theorist to do? One very promising option is to formulate the agency theory within a particular framework that allows for irreducible yet physically grounded properties and that, on some versions, endows objects that possess them with novel causal powers: the emergentist framework.

In the past century, a number of philosophers have pumped and exploited the intuition that ‘the whole is greater than the sum of its parts’ (see Morgan, 1923; Broad, 1960; Alexander, 1966). This is, crudely put, the central tenet of a philosophical doctrine known as ‘emergence’. Contemporary emergentist philosophers typically appeal to the notion of emergence in order to explain the relationship between the mental and the physical, arguing that mental properties are ontologically emergent properties that are generated from the sufficiently complex configuration of the microphysical properties of the brain.

The philosopher Timothy O’Connor attempts to make sense of agent-causation “as an emergent capacity of a fundamentally biological system” (O’Connor, 2000a,

---

6 There are those that explicitly reject the possibility of ontological emergence and endorse only representational emergence that states that emergent properties are conceptually but not ontologically irreducible.
Like all standard agent-causationists, O'Connor eschews event-causation as the model on which our free actions are to be understood, arguing that event-causation does not provide agents with the requisite ability to either act or not act under a certain set of circumstances, and thus does not allow for truly free actions. Rather, O'Connor takes free action to be grounded in an emergent property of purposive agents, that is, agents that have the capacity to represent different courses of action, along with the capacity to entertain certain reasons for or against pursuing each course of action. Given this rational capacity, purposive agents are able to "directly bring about immediate states of intention to act in various ways" (O'Connor, 2000a, p.113). All of this is made possible by a special kind of emergent property that "enables the individual that has it in a certain range of circumstances to freely and directly bring about (or not bring about) any of a range of effects" (O'Connor, 2000b, p.121) - just as our experience of agency indicates that we are able to do.

In order to fully understand the nature of this "volition-enabling" (O'Connor, 1995, p.177) property, we must understand the specific variety of emergence that O'Connor advocates. O'Connor's account of emergence is well worked-out and rich in subtle details. For my purposes, it will suffice to offer a general overview of the basic claims to which O'Connor is committed, which may mean glossing over certain points that, although interesting, are not directly relevant here. O'Connor defines an emergent property as a "nonstructural, natural property that is exemplified by objects or systems that attain the appropriate level and kind of organizational complexity and that exerts downward causation" (O'Connor, 2000b, p.111). There are two key points to note here:
(i) that emergent properties are ‘nonstructural’, and (ii) that they exert ‘downward causation’.

According to O’Connor, a property is nonstructural if “its instantiation does not even partly consist in the instantiation of a plurality of more basic properties” (O’Connor, 2000a, p.108). This feature of emergent properties distinguishes them from properties that are merely ‘resultant’, that is, those properties that are not ‘over and above’ the base properties of the complex whole that possesses them, but are rather ontologically reducible to them. For example, mass is a resultant property, since the mass of any object is just a sum total of the masses of the micro-level particles of which it is made. A corollary of the non-structural nature of emergent properties is that their generation cannot be predicted or explained by appealing to the laws that govern their base properties. In this regard they are entirely *sui generis*, and are often characterized as fundamentally novel types of properties in the ontological furniture of the world. In order for us to see how all this might come together, O’Connor offers the following theoretical scenario:

As a simplifying assumption, suppose there were just one naturally emerging property, P, and that P is instantiated in systems of a high level of complexity, n. Imagine further that physicists used a set of fundamental laws, L, to accurately describe physical processes in all systems whose levels of complexity were lower than n, although the laws failed fully to govern these complex systems of level n. In this scenario, there would be good reason to surmise that we had an emergent property (or properties) at work” (O’Connor, 2000b, p.113).

The term ‘downward causation’ refers, quite simply, to the power of these novel types of properties to produce effects on the lower-levels from which they emerge. There is much more that can and has been said about this concept, for example, some
distinguish between non-reflexive, reflexive, synchronic, and diachronic species of downward causation (see Kim, 1992), associating various problems and advantages with each version, but addressing these issues will take us too far from our present focus, so I will not endeavour to do so here.

I now turn to consider two objections to O’Connor’s general account of emergence. The first one targets the claim that the activity of a complex system bearing emergent properties could not be accurately described using a set of laws that appeals only to the base properties. The trouble with this claim is that it seems perfectly reasonable to suppose that one could describe the new behaviour with the addition of disjuncts to the laws already governing the microproperties. These laws would not have to make reference to any emergent properties, but rather, could be stated entirely in terms of the base properties on which the emergent properties depend.

O’Connor, I think correctly, points out that quite a bit of parsimony would be lost in this approach. The laws used to describe the behaviour of these systems would be “oddly complex” and strangely discontinuous from the laws governing systems of lesser complexity. Thus, O’Connor concludes that this is not a reason to reject his account of emergentism.

Although I agree with O’Connor that adopting disjunctive laws may not be desirable, it is far from clear that positing a completely novel type of property is any more desirable. To suggest that there exists a special kind of property not found anywhere else in nature in order to explain a particular phenomenon seems to me even more theoretically radical than to suppose that in certain situations the natural laws we are working with will be unmanageably complex. If we were to adopt the former strategy
every time we came across a seemingly unexplainable phenomenon in nature we would be hard-pressed to get anywhere in building general theories of how the world works.

The second, and related, objection is that the novel causal powers, which our natural laws are claimed to be inadequate to predict and explain, may not be grounded in emergent properties at all, but rather in the microparticles from which they supposedly emerge. There are two options here: either the microparticles themselves have latent causal features that may only be activated when they are embedded in systems of the requisite complexity, or the entire complex base that is composed of microparticles interacting on various levels itself has the property that endows the system with the causal power (O’Connor, 2000b, p.113).

To the former option, O’Connor replies:

But this won’t do. In the imagined scenario, a microparticle would have different effects in the same type of local situation, depending on the broader context in which that local situation is itself imbedded. What could explain the ‘responsiveness’ of micro-level behavior to macro-level circumstances? (O’Connor, 2000b, p.113).

One way to respond to these remarks is to point out that the very property of generating emergents - supposedly possessed by microparticles - is something that seems to be activated only under certain macro-level circumstances, namely the ones in which there is a sufficiently complex whole (O’Connor, 200b, p.113). Thus O’Connor is equally guilty of supposing that micro-level behaviour might be responsive to macro-level circumstances. But O’Connor retorts that it is not the case that the property of generating emergents is activated only once a certain threshold of complexity is reached, but rather that “[e]ach microparticle is always ‘doing its thing’ (in any context) toward the end of generating an emergent” (O’Connor, 2000b, p.114), it is just that in most
cases, this tendency is "indiscernible" (O'Connor, 2000b, p.114). There is no real way to demonstrate that this is not the case, but it is worth pointing out that one might be equally entitled to say that each microparticle is always contributing to the causing of a certain effect that will not be brought about unless the right threshold degree of complexity is met, and that its seemingly novel causal features are not latent, just indiscernible in any context.

It is also possible, however, to switch gears here and pursue the second option, so that rather than attributing the novel causal powers to microparticles, one could attribute them to the entire base complex. O'Connor objects to this strategy on the grounds that the base complex "does not have an existence of its own" (O'Connor, 2000b, p.113) and therefore no activity of the complex can be additional to the activity of its constitutive parts. But isn't it possible that the complex base has these novel causal powers via the special combination of the causal powers of its constituent parts, rather than in addition to them? This would not entail the claim that the base property had any existence 'over and above' that of its parts.

I will not push further to resolve these matters here, for I am willing to work on the generous assumption that it is at least logically possible that emergent properties, in the sense in which O'Connor describes them, i.e. non-structural and able to exert downward causation, exist and that whether they actually do or not is a matter that may be settled elsewhere. The point of the preceding discussion is merely to highlight some potential challenges that O'Connor's brand of emergentism faces and the fact that the assumption I start with is far from uncontroversial. Nonetheless, I make the assumption
so that I can coherently evaluate O'Connor's volition-enabling property, which is my main interest here.

What is to be made of O'Connor's radical supposition that there exists a unique brand of emergent property, dubbed a 'volition-enabling property' that is able to ground novel types of causal powers in *purposive agents* alone? The causal powers linked to this unique species of emergent property are quite different than the causal powers grounded in ordinary properties. In the latter case, under appropriate circumstances, ordinary properties will contribute to causal powers that will yield a certain effect out of necessity or probability (O'Connor, 1995). For example, a match has certain properties that under the right circumstances, like being struck, there being enough oxygen, etc., will contribute to the causal power to produce the effect of fire. Now the function from the causal power to the effect can be probabilistic, so that, for example, in the exact same circumstances striking the match will result in fire only 70% of the time, or it might be necessary, so that fire is the result all of the time. In contrast, the volition-enabling property grounds causal powers that are neither probabilistic nor necessary, but rather that "*make possible* the direct, purposive bringing about of an effect by the agent that bears it" (O'Connor, 1995, p.177).

So, for example, suppose that given the particular state than an agent’s brain is in, and the possession of the volition-enabling property by the agent, there is a 30% chance that intention A will be formed, a 20% chance that intention B will be formed, and a 50% chance that intention C will be formed. In this situation, the agent’s formation of the intention need not be a function of any of these statistics, but is rather entirely up to the agent in virtue of having the volition-enabling property.
Now according to O’Connor,

... given that there is nothing inconsistent about the emergence of an “ordinary” causal property, having the potential for exercising an irreducible causal influence on the environments in which it is instantiated, it is hard to see just why there could not be a sort of emergent property whose novelty consists in its capacity to enable its possessor directly \[sic\] effect changes at will (within a narrowly limited range, and in appropriate circumstances). (O’Connor, 1995, p.181).

Here O’Connor seems to be suggesting that if one concedes, as I am willing to, that it is possible that emergent properties exist, then there is no reason that one should not also concede that volition-enabling properties do too. Is this a warranted inference? In what follows, I will argue that it is not.

I do not have an argument that demonstrates a priori why it cannot be the case that there exists a property that endows its possessor with the ability to perform an action at will. Indeed, as O’Connor himself admits, it may not be possible to offer such a demonstration given how “thin” of a model he has provided of the functioning of this property (O’Connor, 1995, p.180). I think, however, that despite the lack of demonstrative proof, there are good reasons to doubt the existence of the volition-enabling properties that O’Connor endorses.

First, although O’Connor suggests that given the acceptance of emergent properties in the way he describes them, there is no reason not to also accept the possibility of volition-enabling properties, I believe that the burden of proof rests on O’Connor to provide a positive argument for why we do have reason to accept volition-enabling properties in addition to emergent properties in general. O’Connor owes us this much, since the acceptance of such a property commits one to a special sort of causation that is exclusive to purposive agents and not found anywhere else in nature, even if one...
admits that emergent properties can have novel causal powers of the sort O'Conor supposes, which I have already argued is contentious in itself. Rather than provide us with a positive argument, however, O'Connor merely attempts to whet our intuitive appetites by suggesting that his claims might be

... more plausible when considered in relation to entities such as ourselves, conscious, intelligent agents, capable of representing diverse, sophisticated plans of action for possible implementation and having appetitive attitudes that are efficacious in bringing about a desired alternative (O'Connor, 1995, p.181).

However, simply pointing out that there is a feature of human nature that is in need of explaining does not suffice to make a particular way of explaining it desirable. O'Connor needs to demonstrate why his account is better than an account that supposes that mental causation is just a species of event-causation and is consistent with everything we know about how the world works. Of course, there may be problems inherent to such explanations, but the same applies to those of O'Connor.

That being said, another reason for doubt is that the actual account that O'Connor provides is not terribly convincing or informative itself, since it seems to be suspiciously tailor-made to explain free action. After all, he gives no additional explanation as to how the possession of the volition-enabling property might allow one to exercise causal powers in the radical way that he suggests, he merely stipulates that this is the case.

Furthermore, in formulating the emergentist framework within which the volition-enabling property is posited, O'Connor starts out “on the supposition that qualitative and intentional features of our mental states are emergent” (O'Connor, 2000a, p. 109) and builds his account of emergence upon this assumption, using our common-sense intuitions about the way that mental properties function to inform it. The problem is that
he does not provide any extra-mental examples of emergent properties at work, so there is no independent evidence to support his general account. It, too, seems to be customized to the phenomena that it purports to explain. I think that all these reasons suffice to make O’Connor’s attempt at capturing the experience of agency metaphysically through an appeal to volition-enabling properties dubious.

In this chapter I have considered philosophical attempts, in the form of theories of agent-causation, to remain completely faithful to the experience of agency on a metaphysical level. I have shown that agent-causation is fraught with difficulties stemming from the basic claim that an agent, rather than an event, can be the sole cause of an action. Furthermore, one of the most promising ways to defend this claim, i.e., within an emergentist framework as argued by O’Connor, is suspiciously tailored to the phenomenon in question, thus leaving room to doubt its theoretical robustness.
In the last section I argued that attempts by agent-causationists to vindicate the experience of agency fail to overcome significant metaphysical obstacles, and are thus left standing on shaky ground. Does this mean that our experience of agency is not indicative of our ability to cause our own actions, if not as substance wholes, at least in virtue of our intentions? In the past twenty-five years some attacks on the experience of agency on the grounds that it is illusory have emerged from within the trenches of cognitive science. These attempts to dismiss the experience of agency as smoke and mirrors are somewhat novel compared to similar thrusts made by philosophers, in that they rely on evidence from cognitive neuroscience and cognitive psychology to demonstrate their validity. In this section I will examine the two most aggressive campaigns to demonstrate the illusory nature of the experience of agency; one launched by the psychologist Daniel Wegner and the other by the neuroscientist Benjamin Libet. I will argue that both fail to show that our sense of agency is illusory.

In his recent book, The Illusion of Conscious Will (2002), Wegner sets out to show that in some way or other, the feeling of doing is just an intricate illusion. It will be useful to clarify the sense in which Wegner takes the experience of agency to be illusory, as there are many different forms that this thesis could take. There is evidence to support the claim that Wegner, at different times, can be read with each of the following two views in mind: 1. Experiential epiphenomenalism: The experience of agency is illusory in
the sense that it does not actually cause our actions, and 2. Intentional epiphenomenalism: The experience of agency is illusory in the sense that our intentions do not actually cause our actions.

First, I will consider the view that the experience of agency is illusory because it does not cause our voluntary actions. Wegner evidently has this claim in mind when he states firmly that, “… conscious experiences of will do not cause human actions” (Wegner, 2002, p.318). The picture being denied here is one in which the experience of agency itself is causally efficacious, rather than a mental event that it represents. John Searle (1983) defends this view in the following passage:

… in the case of the experience of acting, the Intentional component has the world-to-mind direction of fit. If I have this experience but the event doesn’t occur we say such things as that I failed to raise my arm, and that I tried to raise my arm but did not succeed. And the direction of causation is from the experience of acting to the event. Where the Intentional content is satisfied, that is, where I actually succeed in raising my arm, the experience of acting causes the arm to go up. If it didn’t cause the arm to go up, but something else did, I didn’t raise my arm: it just went up for some other reason (Searle, 1983, p.88).

What could it mean to say that an experience of acting is the cause of an action? Claiming this seems to commit one to the puzzling view that an experience can be the cause of the content of that experience. To see why this view is so strange, we might analogously consider the experiences of pain and visual perception. We would not want to say that one’s experience of pain is the cause of that pain, nor that one’s experience of red is the cause of the red. This is not to say that no experience can be causally efficacious. Indeed, one might reasonably argue that an experience of pain at the hands of a hot stove might cause someone to avoid hot stoves in the future. The difference is in the timing. To say that the experience of acting causes one’s arm to go up is to say that at
the same time that the experience indicates that one’s arm is already going up, it is causing it to go up, and it is this synchronicity that is baffling. Moreover, it is not even clear why it should be beneficial to hold this view, since it would seem to leave no room for the types of entities we normally take to be the causes of our actions, i.e. selves, mental events, etc.

Worthy of more attention, I think, is Wegner’s endorsement of intentional epiphenomenalism. Wegner explicitly espouses epiphenomenalism with regard to intentions, relegating them to mere previews of what will be done and excluding them from the causal picture entirely:

> [t]he co-occurrence of thought and action may happen because thoughts are normally thrust into mind as previews of what will be done... [i]ntentions, in this analysis, are to action what turn signals are to the movements of motor vehicles. They do not cause the movements, they preview them (Wegner, 2004, p.656).

There are two issues lurking here, one having to do with the content of the experience of agency and the other having to do with the extent to which that content maps onto the real world. In order for the claim that our experience of agency misrepresents our intentions as the causes of our actions to be adequately defended, it must be shown that: (i) the experience of agency is an experience of our intentions causing our actions, and (ii) our intentions are not the causes of our actions. For now we will assume the truth of the claim about the content of the experience of agency and focus instead on the question of whether or not Wegner adequately demonstrates that our intentions are not the causes of our actions.

---

1 I should note here how Wegner and I differ with regard to our views on experiential epiphenomenalism. I agree with Wegner that the experience of acting is not the cause of our actions, but we hold this claim for different reasons, for whereas I take the claim that the experience of acting is the cause of our actions to be incoherent from the start, Wegner accepts it as a plausible hypothesis and then rejects it.
The first thing to note is Wegner’s ambiguous stance on the metaphysical status of mental events. Although he never tires of claiming that our intentions have no causal powers, he neglects to elaborate on any other features of our intentions, most importantly their relation, if any, to events in the brain. This elaboration is crucial if Wegner’s account is to be intelligible, for if, like most psychologists, Wegner advocates some variation of physicalism, then his conviction that intentions are epiphenomenal is truly puzzling, since intentions are just types of mental events, which under physicalism must in some sense be physical events, the causal powers of which are postulated without problem. Moreover, nowhere does Wegner offer an explanation detailing what he takes to be the ‘actual’ causes of our actions, if not standard mental events like intentions. The only clarification he offers in this regard are some cursory remarks to the effect that “unconscious” events are at work, causing both our conscious thoughts and our actions, but he does not venture to describe the nature of these events any further. The withholding of fundamental details such as these makes it difficult to evaluate Wegner’s arguments, but I will do my best to do so on the assumption that the claim that our intentions are epiphenomenal is itself coherent and evaluable given certain unspecified metaphysical commitments.

In defence of the view that our intentions are epiphenomenal, Wegner formulates the Theory of Apparent Mental Causation, which lays down three basic principles, the holding of which is supposedly sufficient to generate the experience of agency, whether or not we are actually acting as agents at the time: (i) the priority principle, (ii) the consistency principle, and (iii) the exclusivity principle. According to the priority principle, a conscious thought must be had prior to action. For example, if I raise my left
arm and have a conscious thought about raising my left arm only after doing so, then I will not experience a sense of agency for this movement. The consistency principle states that the conscious thought must be consistent with the action. For example, if I have a conscious thought about raising my left arm and my right arm goes up instead, I will not experience a sense of agency for this movement. Finally, the exclusivity principle requires that the conscious thought must appear exclusive of salient alternative causes to the action. For example, if someone else lifts up my arm while I am simultaneously having a conscious thought about raising it, then I will not experience a sense of agency for this movement. In sum, according to the Theory of Apparent Mental Causation, “for the perception of apparent mental causation, the thought should occur before the action, be consistent with the action, and not be accompanied by other potential causes” (Wegner, 2004, p.656).

In support of this theory, Wegner discusses two types of cases in which there is a disconnect between the sense of agency and the activity of the agent: automatisms and illusions of control. The former are situations in which one lacks a sense of agency for things one is, in fact, doing and the latter are situations in which one has a sense of agency for things one is not doing. In order for Wegner’s theory to be validated, it must be the case that for every automatism, one of the three principles does not hold, and for every illusion of control, all three of the principles hold.

Wegner draws on a variety of sources, sometimes dubious, e.g. Ouija boards and table-turning, in order to demonstrate the possibility of automatisms. The most intriguing of these is known as Anarchic Hand Syndrome; a neuropsychological disorder associated with damage to the frontal lobe that leaves one feeling divorced from the movements of
one's own hand. Those suffering from this disorder deny that they have control over the activities of their hand, often attributing them to an outside source. The movements in question are not subtle either; in fact, their seemingly intentional qualities are what makes this particular disorder so remarkable. Examples of the anarchic hand’s ‘actions’ include unbuttoning shirts, making moves in checkers, and in one case, even grabbing the hand owner’s throat (Wegner, 2002).

Although incidents of Anarchic Hand Syndrome do demonstrate that it is possible to lack a sense of agency for seemingly intentional actions, they do not, nor do automatisms in general, help support the view that the experience of agency is illusory in virtue of misrepresenting the world to us, since it is altogether absent from these situations. What Wegner needs to show is that in all the cases in which we do have an experience of agency for our actions, it falsely represents our intentions as the causes of our actions. As such, cases in which we do not have a sense of agency for bodily movements are not relevant to his epiphenomenalist thesis.

The existence of illusions of control, on the other hand, would provide crucial support for such a thesis, and we should thus examine Wegner’s case for them carefully. In order to prove that it is possible to have an experience of agency for an action that one has not performed, Wegner describes a personal anecdote in which he is shopping at a toy store with his family and experiences a supposed illusion of control:

... I eased up to a video game display and started fiddling with the joystick. A little monkey on the screen was eagerly hopping over barrels as they rolled toward him, and I got quite involved moving him along and making him hop, until the phrase ‘Start Game’ popped into view... I had been fiddling for nothing, the victim of an illusion of control.” (Wegner, 2002, p.10).
Is this a genuine example of an illusion of control? Seemingly not, for it is not the case that Wegner's sense of agency mistakenly accompanies his own actions, but rather the effects of those actions. He is, of course, experiencing a sense of agency with regard to controlling the virtual monkey, but only in virtue of having an accurate sense of agency regarding his moving the joystick. One way to highlight the mistake that Wegner is making in supposing that this is a valid example of an illusion of control is to consider an analogous attempt to say that someone who does not feel a sense of agency for the movements of the monkey on the screen in virtue of his joystick fiddling, where the causal connection does hold, is the victim of an automatism. Clearly this is misguided and neglecting the main point, which is the agent's experience of his own movements, not the effects they cause.

Another attempt by Wegner to offer an example of an illusion of control is equally unsuccessful:

The illusion that one has done something that one has not really done can also be produced through brute social influence, as illustrated in a study by Kassin and Kiechel (1996). These researchers falsely accused a series of participants in a laboratory reaction time task of damaging a computer by pressing the wrong key. All the participants were truly innocent and initially denied the charge, showing that they didn't really experience damaging the computer. However, they were led later to remember having done it. A confederate of the experimenters claimed afterwards that she saw the participant hit the key or did not see the participant hit the key. Those whose "crime" was ostensibly witnessed became more likely to sign a confession... internalize guilt for the event, and even confabulate details in memory consistent with that belief - but only when the reaction time task was so fast that it made their error seem likely. (Wegner, 2002, p.10)
Again, this does not seem to be a genuine example of an illusion of control. The subjects in this experiment denied that they experienced a sense of agency for pressing the wrong key, which is the only relevant factor in this scenario. Whether or not they were successfully convinced later on, through whatever means of social coercion, that they had pressed the wrong key, although interesting for social psychology, makes no difference to the question of whether or not they had the sense of pressing the wrong key at the time that they ostensibly pressed it, and this is our primary concern.

So far we have considered two failed attempts at offering an appropriate example of an illusion of control. Wegner does, however, offer one other example, meant to support the priority principle of the Theory of Apparent Mental Causation, that may also provide support for the claim that the experience of agency is sometimes generated in cases in which no action has occurred: the ‘I Spy’ study. In this study, Wegner and Wheatley (1999) conducted an experiment inspired by the common Ouija board in which they aimed to discover whether or not participants would feel that they had voluntarily moved a pointer on a computer screen if they had a prior thought about the movement.

Participants and confederates were paired up, and both were seated facing each other with a small board on top of a computer mouse in between them. Beside the two was a computer screen showing several small objects. The participants were instructed to move the mouse together by putting their fingers on the board and moving them in slow sweeping circles. They were to stop the pointer on a particular object approximately every thirty seconds during a 10-second interval of music played over headphones that they were asked to wear, and then rate the level of intention that they had for each of these stops on a special scale, which would later be converted to a percentage.
During each trial, the participant and confederate would hear the same word over their headphones, although they were told that the words they would be hearing were different. The words were meant to serve as primes for thoughts about certain objects on the screen. On certain trials, the confederate would force stops on whatever object matched the word they both heard whereas on other trials the subject was allowed to stop on whatever object he or she pleased with no interference from the confederate. The results showed that on the 'forced' stop trials, the participants perceived their actions as more intentional in the situations in which the word was spoken closer to, rather than further away from, the time at which the confederate stopped on the object. Perhaps even more interestingly, Wegner and Wheatley (1999) found that the average level of intention reported for the unforced stops was 56% - around the same level as that reported for the forced stops in which the priming word occurred very close in time to the actual stop.

Wegner and Wheatley (1999) controlled for the possibility that the patients actually did intend to stop on the objects during the forced trials by comparing the mean distance between the stop and the object on all unforced trials in which the word named an object on the screen vs. unforced trials in which the word did not name an object on the screen. It can only be assumed that on the trials in which the word did not name an object on the screen the subjects were asked which object they were intending to stop on. They found no significant differences between the mean distances, allowing them to conclude that the participants were not contributing to the stop on the named object during the forced trials. Thus, they concluded that "the experience of will can be created by the manipulation of thought and action in accord with the principle of priority, and this
experience can occur even when the person’s thought cannot have created the action” (Wegner, 2002, p.78).

Should we accept the ‘I Spy’ experiment as an example of an illusion of control? And if so, should we conclude from it that the experience of agency misrepresents our intentions as the causes of our actions? Two things should give us pause before we answer either of these questions in the affirmative: (i) the level of intention reported by the subjects for the stops, and (ii) our attitudes towards visual illusions.

The fact that the level of intention for both the forced and unforced stop trials was roughly the same cannot alone indicate that the subjects had an experience of agency for stopping the pointer on the target object. The low level of intention reported by the subjects during the forced trials cannot be ignored. The average level of intention reported was around halfway between the “I allowed the stop to happen” and “I intended to make the stop” levels (Wegner, 2002), indicating that the subjects were hesitant, for whatever reason, to say that they had fully intended to make the stop. My suggestion is that, just like in Wegner’s misguided example of experiencing an illusion of control while playing a video game during its demo, the subjects were not so much confused about their own movements, but rather about what effects those movements were having on the pointer on the screen, given that the confederate was supposedly moving the mouse along with them. After all, one would be very surprised if the subjects reported the same medium levels of intention for their actual finger movements. The effect Wegner created was indeed exploiting certain inference mechanisms in the brain, but only the ones that we use to infer causal relationships in the environment, not in ourselves. The subjects
may have falsely attributed the movement of the pointer on the screen to their finger movements, but they did not also falsely attribute their finger movements to themselves.

We have as much of a reason to believe that our experience of agency always misrepresents the world to us as we do to believe that this is the case for our visual experience as well. It is common to experience visual illusions from time to time, for example when one perceives a straight stick to be bent in water, or when one is suffering from a hallucination. It would, however, be unreasonable to view these instances as indicative of more ubiquitous misrepresentation by our visual system. Likewise, we should not suppose that illusions of control, if indeed a concrete example can be offered, imply that our sense of agency never accurately represents the world. The fact that there are flaws in the system under certain conditions does not mean that the system never works. Thus, although Wegner's work on the experience of agency exposes some issues worthy of further examination and raises interesting questions surrounding the content of the experience, the conditions under which it is generated, and its functional role, his specific claims regarding the veridicality of the experience of agency are, in my opinion, largely unfounded and should be dismissed until they can be more convincingly defended.

I turn now to another proponent of the view that our experience of agency is illusory in some sense: the neuroscientist Benjamin Libet. Whereas Wegner draws support for his claims from a variety of sources, Libet's claims are almost entirely founded upon one single experiment that he and his colleagues conducted in the early 1980's, which purportedly showed that our conscious decisions to initiate an action are always preceded by brain events that themselves signal the initiation of voluntary action.
In this experiment, subjects were seated facing a visible analogue clock around which a dot would revolve once every 2.56 seconds. They were asked to keep a close watch on the position of the dot on the clock and to perform a quick flick of the wrist whenever they felt like doing so, without any premeditation. After a few seconds, the subjects were instructed to report, to the best of their abilities, the position of the dot on the clock face at the time that which: a) they were aware of their ‘intention’ to act, an event which Libet and his colleagues labelled the ‘W judgment’, or b) they were aware of actually moving, an event which Libet and his colleagues labelled the ‘M judgment’, or c) they were aware of a sensation produced by the external application of a stimulus pulse to the back of their hands, which Libet and his colleagues labelled the ‘S judgment’. The S judgment was used as an indication of the subjects’ accuracy in making the W and M judgments, and to control for the well-acknowledged difficulty in making cross-modal judgments of the sort required of them (Libet et al., 1983). These three types of judgments were never made on the same trial. Rather, the experimenters held series of trials – forty trials in each – consisting solely of one type of judgment. For each of the W and M trials, strategically placed electrodes were used to measure the onset of the Readiness Potential, or RP, which is a brain event that regularly precedes self-initiated acts by up to one second, and an electromyogram (EMG) to measure the exact time at which the voluntary motor act was initiated.

In my discussion of this experiment, I will only be interested in the onset of the RP in relation to the time reported by the subjects in their W judgments, since it is from these measurements that conclusions were drawn regarding the veridicality of the experience of agency. It turns out that, on average, the onset of the RP in the brain was
measured to precede the time indicated in the W judgment by at least 150ms and by an average of 350ms, even after accounting for the subjects' bias using their S judgments. In addition, the subjects reported being aware of the decision to flick their wrists an average of 150ms before the onset of the movement as measured by the EMG. In all of the trials in question, the subjects reported being unaware of any pre-planning of when to move and indicated that their movements were "spontaneous" and "capricious" (Libet et al., 1983, p. 632). From these results, Libet and his colleagues concluded that, "... cerebral initiation of a spontaneous, freely voluntary act can begin unconsciously, that is, before there is any (at least recallable) subjective awareness that a 'decision' to act has already been initiated cerebrally. This introduces certain constraints on the potentiality for conscious initiation and control of voluntary acts" (Libet et al., 1983, p.623).

What significance do Libet's results and the conclusions he draws from them have for the phenomenology of agency? The implication seems to be that our experience of agency, which includes the experience that I am the cause – and thus the initiator - of my actions, is illusory in the sense that it is not me at all, but rather some 'unconscious' cerebral events that are doing all the work of initiating my flick of the wrist. After all, the time at which I have the experience of causing my action is, by neuroscientific standards, long after the causal chain of events leading up to the action has already been initiated in the brain without any of my input. Indeed, the experience of agency cannot be an accurate indicator of my agency if it does not correspond with the actual initiation of my action, and thus we are back on the familiar path of viewing the experience of agency as an illusion.
Libet’s own interpretation of his results is that in the 150ms interval between the reported time at which there is an awareness of the intention to act and the time at which the act itself occurs, the “conscious function might affect the final outcome of the volitional process” (Libet, 1999, p.49). Libet suggests that although we may not have the power to initiate our voluntary actions, we have the power to exercise a veto over them once they are already underway, thereby preventing them from being executed, if we so please. Thus, the experience we have of initiating our actions is really the experience of allowing a previously initiated action to go through, where ultimately it is ‘up to us’ whether or not it does. Daniel Dennett (1991) offers a nice summary of Libet’s interpretation:

[we are not quite “out of the loop” (as they say in the White House), but since our access to information is thus delayed, the most we can do is intervene with last-moment “vetoes” or “triggers.” Downtown from (unconscious) Command Headquarters, I take no real initiative, am never on the birth of a project, but do exercise a modicum of executive modulation of the formulated policies streaming through my office (Dennett, 1991, p.164).]

This modest preservation of some control over our actions does not help us defend against the conclusion that the experience of agency is illusory, however, for the experience of agency includes the experience of being the cause and initiator of my actions, where this entails being the author of them, and not just the final editor.

One way to resist the troublesome implications of Libet’s interpretation is to endorse what I will call the ‘Volitional Interpretation’ of the results. The Volitional Interpretation is supported by Jing Zhu (2003) and hinges on two important features of the experiment: (i) the general prior intention to act in compliance with the instructions of the experiment, and (ii) the object of the subjects’ introspection. First, Libet largely
ignores the fact that prior to flicking their wrists, their intentions to flick their wrists, and the onset of the RP, each subject forms the intention to go along with the experiment and perform a flick of the wrist at an unspecified time during the trial while attempting to introspect on their decision to do so. The suspicion is that this general prior intention to act may have an effect on the act itself, thereby precluding it from being truly 'spontaneous' and 'capricious' as required (Zhu, 2003). After all, the subjects have practiced the movement in question many times and are aware of exactly what they are going to do. The only variable that remains unknown is the exact time at which the execution will occur, and even this can be narrowed down to a rough time interval.

Second, the correct term for the object of the subjects' introspection for the W judgment is unclear. Libet et al. describe the object of introspection as a 'wanting', 'urge', 'decision' or 'intention' to act, noting that subjects most often preferred the terms 'wanting' or 'urge' (Libet et al., 1983, p.627). I will assume here that these two terms can be used synonymously, and more specifically, that they can both be accurately characterized as 'sudden desires'. That being said, there are significant differences between sudden desires to act and intentions to act that cannot be ignored in this context. For one thing, while intentions to act are generally formed for reasons, which may in turn consist of certain beliefs and desires, sudden desires to act are generally not grounded in rationality. For example, I may I intend to call my mother today because I haven’t spoken to her in over a week, whereas I may have the sudden desire to pull a fire alarm upon walking past it, despite having no reason to do so. Furthermore, sudden desires often spring up 'out of nowhere' and we might not 'see them coming' until it is almost
too late to stop the act that they prescribe. Intentions, on the other hand, are not spontaneous in this way, as we play a direct role in their formation and execution.

Combining the claims that the subjects had a general prior intention to flick their wrists and that what they eventually became aware of upon introspection was a sudden desire, begun unconsciously, to flick their wrists, Zhu offers an interpretation of the results that preserves veridicality of the experience of agency while accounting for the early onset of the RP:

The subjects in the Libet's [sic] experiment were instructed to look for feelings of 'wanting to move'. A mechanism of selective attention was thereby activated to detect the urge to move. They were in a state of readiness to move and expected a signal, which then triggered the predefined well-learnt motor act. When such an urge is detected, the [supplementary motor area] participates in initiating conscious spontaneous movements, which correspond to the onset of a typical RP. As a result of selective attention, the subjects could perceive a normally unconscious process that was later reflected in the subjects' introspective awareness (Zhu, 2003, p.71).

There are two important things to note about this interpretation. First, the subjects aren't merely passively introspecting on their urges, but rather they "intentionally make certain when they [are] strongest, so that the designed 'spontaneous' actions [can] occur" (Zhu, 2003a, p.70). Second, the way the subjects select a certain urge to be the one that leads to the action is by "exerting a volition" (Zhu, 2003, p.72). So the picture we have here is of the subject playing the ever clever orchestrator of brain events; monitoring various urges that have their source in the original intention to go along with the study as they float by in the stream of consciousness; swiftly plucking out the strongest one and finally performing a volition to produce the flick of the wrist, all the while passing effortlessly through the boundaries between consciousness and unconsciousness.
Although this is one interpretation according to which our agency is preserved, upon closer examination, Zhu (2003) relies on some rather bold assumptions, the theoretical cost of which might not be worth the benefits. For one thing, Zhu’s suggestion that we make use of special metaphysical tools like ‘volitions’ to control the movements of the body is questionable. In support of this hypothesis, Zhu appeals to a study that sought to examine the difference between the RP accompanying “unconscious spontaneous movements” and conscious, voluntary movements such as those in the Libet experiment. The results of the study indicated that the RP for conscious, voluntary movements is much stronger than it is for the unconscious ones:

There are two remarkable differences between the two types of RP. First, RPs due to unconscious bodily movements had much smaller amplitudes than RPs associated with conscious movements (Libet situation), which suggests the greater amplitude of RP may be associated with conscious effort. Second, but more striking, the scalp distribution of RPs correlated with conscious spontaneous motor acts was different from that measure from unconditionally performed movements. [...] How can we account for the difference between the two kinds of internally initiated movements? A plausible answer is that, in the Libet situation, people exert volitions to initiate intentional actions, whereas for unconscious spontaneous movements [...] no volition is involved (Zhu, 2003a, p.72).

Although I agree that the difference between the RP in the cases of ‘conscious’ and ‘unconscious’ movements is certainly in need of explaining, I do not see how positing the involvement of volition in one case and not the other is helpful. First of all, the concept of volition itself has enjoyed a controversial history in philosophy, and is perhaps less frequently used these days for good reason. One of the staunchest advocates of volition was John Locke, who claimed that “[v]olition, ‘tis plain, tis an Act of the Mind knowingly exerting that Dominion it takes it self to have over any part of the Man, by employing it in, or withholding it from any particular Action” (Book II, XXI, §15).
The problem with volition described as such, i.e. as a special exertion of the mind that serves as a catalyst for voluntary but not involuntary actions, is that it seems less like a good explanation of voluntary action and more like it is in need of a good explanation itself. After all, it is not clear how volitions are supposed to relate to the rest of our folk psychological repertoire that includes beliefs, desires, and intentions, nor where they lie in the mental state taxonomy, which consists of intentional states and sensations. Gilbert Ryle highlights the mystery surrounding the ‘doctrine of volition’ in his quizzical remarks:

By what sorts of predicates should [volitions] be described? Can they be sudden or gradual, strong or weak, difficult or easy, enjoyable or disagreeable? Can they be accelerated, decelerated, interrupted, or suspended? Can people be efficient or inefficient at them? Can we take lessons in executing them? Are they fatiguing or distracting? Can I do two or seven of them synchronously? Can I remember executing them? Can I execute them, while thinking of other things, or while dreaming? Can they become habitual? Can I forget how to do them? Can I mistakenly believe that I have executed one, when I have not, or that I have not executed one, when I have? At which moment was the boy going through a volition to take the high dive? When he set foot on the ladder? When he took his first deep breath? When he counted off ‘One, two, three – Go’, but did not go? Very, very shortly before he sprang? What would his own answer be to those questions?

Now, there may be accounts of volition that are not committed to such a puzzling definition of the concept, indeed, the literature on volition is extensive and far outside the scope of this paper, but insofar as a volition is taken to be an act of the mind that precedes voluntary and not involuntary movements, as it appears to be in Zhu’s Volitional Interpretation, I admit to sharing Ryle’s confusion.\footnote{Although I do not wish to go all the way with him in his reduction of all mental talk to talk about behaviour and dispositions to behave.}
The difference between the RPs in the movements of which the subjects were aware versus those of which they were not can be readily explained by appealing to less radical mental processes than volitional ones. It is, after all, not surprising that different cognitive tasks, e.g. doing mental calculation versus introspecting on a decision to act, should yield different patterns of activity in the brain, nor that different structures in the brain are responsible for producing voluntary versus involuntary movements. That being said, it seems entirely plausible that the reason that there exists a difference in the RP activity is simply because intentional actions require more cognitive resources than unintentional ones, including planning, attention, motor control, etc.

Another problem with the Volitional Interpretation of Libet’s study is that it posits a homunculus self that sits in the ‘seat of consciousness’ calling all the shots in the brain of which it is in charge. It is this self that calls urges to act up from the unconscious into the conscious arena where it resides and then does what it needs to do in order to have them culminate in a flick of the wrist. The philosopher Daniel Dennett is a strong opponent of this view of consciousness and the self, rejecting it as a relic of Cartesian dualism that ought to be disposed of in favour of a more empirically informed model of consciousness that is compatible with what present day cognitive science has revealed about the workings of the brain. On Dennett’s model, the whole brain is a buzzing melee of synchronous activity, quickly processing stimuli in various parts and on various levels of awareness. There is no conscious workbench on which everything takes place, but rather conscious and unconscious blur together in such away that any myopic focus on conscious processes at the exclusion of all others for the grounding of control and
decision-making is unwarranted. All this entails that the time t at which subjects in Libet’s experiment report being aware of an intention to act loses its significance as well:

When we remove the Cartesian bottleneck, and with it the commitment to the ideal of the mythic time t, the instant when the conscious decision happens, Libet’s discovery of a 100-millisecond veto window evaporates. Then we can see that our free will, like all our other mental powers, has to be smeared out over time, not measured at instants. Once you distribute the work done by the homunculus (in this case, decision-making, clock-watching, and decision-simultaneity-judging) in both space and time in the brain, you have to distribute the moral agency around as well. You are not out of the loop; you are the loop. You are that large. You are not an extensionless point. What you do and what you are incorporates all these things that happen and is not something separate from them. Once you can see yourself from that perspective, you can dismiss the heretofore compelling concept of a mental activity that is unconsciously begun and then only later “enters consciousness” (where you are eagerly waiting to get access to it). This is an illusion since many of the reactions you have to that mental activity are initiated at the earlier time – your “hands” reach that far, in time and space (Dennett, 2003, p.242).

Why do some view this picture as unpalatable? The main worry is that it excludes the possibility that ‘everything comes together’ in consciousness in such a way that all mental processes related to our goals as agents are under our ‘direct conscious control’. It is an outright rejection of the Cartesian conviction that we are infallible as to the contents of our minds, adding layers to the mind of which we are unable to plumb the depths. This inaccessibility makes many uneasy, as it suggests that much of what we do is strongly influenced by processes that we are not privy to, and thus our actions are not ‘up to us’ in the sense that we normally suppose. If Dennett is right though, our commonsense model is neither realistic nor desirable, and we should pursue all efforts to wean ourselves of off it. We do not need to take the radical step of adopting Dennett’s model of consciousness, but it does expose some potential problems with the Volitional Interpretation, namely its commitment to a sort of homunculus self.
Having canvassed the potential weaknesses of the Volitional interpretation, we are now in a position to formulate a satisfactory interpretation of Libet’s results that is free from troubling metaphysical commitments and yet resists the conclusion that our experience of agency is illusory. I will now put forward what I will call the ‘Autopilot Interpretation’, which makes use of both the fact that subjects form a general prior intention to flick their wrists according to the instructions of the experiment, and the fact that they describe their actual wrist flick as an ‘urge’. The wrist flick here would be akin to the situation in which one prepares to rip off a band-aid. One knows exactly what to do and how to do it; it is just a matter of picking the particular time at which one will do it. Upon ripping the band-aid off, one might even be surprised at having done so and not quite realize what has happened until it has been removed. It seems that in a situation like this, it takes us a while to ‘catch up’ with what we have done, much like in the case of unthinkingly pulling a hand away from a hot stove, where the awareness of what we have done is slightly delayed.

I wish to suggest that the wrist flick in Libet’s studies is halfway between a hot-stove reflex and a full-fledged intentional action. The process of performing the act is largely structured by the prior intention to perform it, which adds a layer of executive control not found in regular reflexes. The act itself, however, is very much like a reflex. It is a very simple, basic action for which we really need not commit many cognitive resources, like we might for more complex actions. Indeed, Libet’s study says nothing about the performance of non-simple motor acts, which we tend to more readily associate with our agency. As such, without postulating volitions or homunculus selves, I conclude
that Libet’s results do not warrant the assumption that we do not initiate our actions in the way that we normally suppose.

In this section, I have attempted to show that two of the strongest attempts to expose the experience of agency as an illusion are benign. Not only does Wegner fail to offer a convincing example of an illusion of control, but he does not recognize that even if he were able to provide a concrete example to that effect, it would merely demonstrate that the experience of agency can, just like our visual experience, sometimes misinform us, not that it always does. Libet, on the other hand, is too hasty in the conclusions he draws, neglecting the importance of the prior intention to act in accordance with the instructions of the experiment and the simplicity of the act in question.
A SKETCH OF THE PHENOMENOLOGY

So far my focus on the experience of agency has been motivated by the question of whether or not it accurately represents us as the causes of our actions. I have attempted to show that neither an overly enthusiastic endorsement of its veridicality, in the form of theories of agent-causation, nor an overly skeptical rejection of it, as espoused by Libet and Wegner, is warranted. I now turn to take a closer look at the phenomenological components of the experience and what they can tell us about its relation to the self.

I begin my discussion by looking more closely at Anarchic Hand Syndrome; the aforementioned condition in which the experience of agency breaks down, leaving those afflicted by it feeling divorced from the movements of one of their hands. Although their hands may perform seemingly intentional, goal-oriented movements, patients with Anarchic Hand Syndrome deny having an experience of agency for them. The psychologist Sergio Della Salla (2005) relates a particularly striking piece of anecdotal evidence for this disorder:

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 left-over fish-bones and put them into her mouth (Della Salla et al., 1994). 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 place and

---

9 The term ‘Alien Hand Syndrome’ is often used synonymously with ‘Anarchic Hand Syndrome’ in the literature (Della Salla, 2005, p.609) but this is a mistake. Alien Hand Syndrome refers to a disorder in which one does not feel that one’s hand is one’s own, but this experience is unrelated to any actions the hand might perform, which is the pivotal aspect of Anarchic Hand Syndrome (Marcel, 2003, p.76).
as a result of the fighting the dessert dropped to the floor. She apologized 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' (Della Salla, 2005, p.606).

There seem to be four distinct features missing from the experience that Mrs. GP and other Anarchic Hand patients have while their deviant hands perform unwanted movements. First of all, there is some sense in which the patients fail to experience the movements of their hands as their own. Second, patients do not experience control over the affected hand’s movements, often struggling or protesting to get it to stop whatever it is doing. In some cases, patients are left with no choice but to sit on the offending hand or pin it down with the unaffected hand (Marcel, 2003, p.77). Third, patients deny intending to perform any of the movements that the hand performs. Fourth, the patients are aware of the movements of the hand only via proprioception and visual or tactile perception rather than 'from the inside' (Marcel, 2003, p.79). The reactions of these patients to their anarchic hand’s movements and the features of their experience that their reports indicate is missing, imply that the sense of agency normally consists of: (i) a sense of ownership, (ii) a sense of control, (iii) awareness of intention, and (iv) awareness of bodily movement ‘from the inside’. I shall now consider each of these in turn.

The most striking property of cases of Anarchic Hand Syndrome is the absence of a sense of ownership for the activities of the deviant hand. Indeed, the first question that this raises is the status of the movements of the anarchic hand given this feeling of alienation from them. Most often, they are referred to as actions, but is this label really appropriate? It is true that the movements of the anarchic hand are highly purposive in their nature, as evidenced by their carrying out complex movements that are clearly directed towards a certain goal, e.g. unbuttoning a shirt, making a move in checkers.
Thus, it certainly *seems* like they are actions, betraying conscious intentions on the part of the agent, but it seems just as certain, given the adamant claims of the agent to the contrary, that they are not.\textsuperscript{10}

All these issues highlight the need for a clarification, i.e. that the sense of ownership is not grounded solely in the awareness of one’s bodily movements. This becomes clear when one considers that bodily movements like knee-jerks or nervous tics are experienced as one’s own only in the sense that they are recognized as movements coming from one’s own body. A similar restriction exists in cases of Anarchic Hand Syndrome. As one man afflicted with the disorder put it, “of course I know that I am doing it\textsuperscript{11}. It just doesn’t feel like me” (Marcel, 2003, p.79). Thus, the experience of ownership might be more accurately described as the experience of one’s own bodily movements as one’s own *actions*, and the question now becomes what needs to be the case in order for *this* experience to be generated? I propose that what is missing is a *sense of control* over the bodily movements. This sense of control might vary in degree. Consider, for example, the degree of control you experience when using your dominant hand versus your non-dominant hand. But it is reasonable to believe that it must at least be present in some minimal form in order for a bodily movement to be experienced as an *action*.

\textsuperscript{10} Some, like John Campbell (2003) deny that a movement can be purposive yet unintended: “There is only one person around who could be the agent of the action, so if the movements of the hand are purposeful they simply must be intended by the agent” (Campbell, 2003, p.162). I think Campbell is mistaken, as a purposive movement, i.e. a movement with a purpose, might be accurately defined as a movement with a goal and this does not require that it be intended. For example, jumping out of the way of an oncoming car before one is even consciously aware of what is happening is goal-directed, i.e. it is purposive, but it is not an intentional action.

\textsuperscript{11} It should be noted that this man most likely means that he knows that he is doing it in the sense that the action originates from his body, rather than in the intentional sense.
In order to have the experience of control over one’s bodily movement, one must at least have the experience of being able to initiate the bodily movement whenever one pleases. Indeed, it is the possibility that one might lack this ability that makes us uneasy when faced with the conclusions drawn from the Libet et al. (1983) experiments. The worry there is precisely that if I am not capable of initiating my bodily movements, then I do not have the requisite control over them. This often leads to broader worries about free will and moral responsibility, but we need not subscribe to these in order to appreciate the point that an experienced inability to initiate a bodily movement will deny one a sense of control over that bodily movement. The Libet experiments also support the claim that one must feel that one is able to terminate an already initiated action in order to have a sense of control over it. This is exactly why Libet’s suggestion that we have a veto power over our actions, that is, we at least have the power to stop them once they have started, is supposed to reassure us that we still have some sort of control over our wrist flicks.

Experiencing the ability to initiate/terminate an action does not appear to be sufficient for a full sense of control though. Consider as an example the throwing of a dart. If I experience the ability to initiate this action and the ability to stop my arm at some point during its trajectory, but no other control of the movements of my arm in between, surely I will feel that my control over this action has been thwarted. What is missing here? A useful analogy might be that of operating a motor vehicle. Not only are gas and brake pedals required in order to have proper control over a car, but a steering wheel as well. So what is missing from the account of control including only the ability to initiate and terminate an action is the ability to guide one’s actions. This type of
'guidance control' might be realized in a sort of feedback loop that allows one to correct one's movements by comparing a representation of the intended action to that of the actual state of the body. The degree of control we experience may be a function of how close we are to executing one of our intentions. For example, if someone is trying to draw a perfectly straight line, they will experience a sense of control insofar as they experience themselves being able to initiate or terminate this action and they will feel more or less control depending on how well they feel they are able to guide the movements of their hand.

One might protest that in cases of Obsessive Compulsive Disorder (OCD), in which a patient experiences strong urges to perform certain actions, often in a very specific way, the patient experiences her actions as her own, but lacks any control over them since she cannot resist the urges. Therefore, the sense of control is not necessary for the sense of ownership, as I have suggested. There are two ways to handle this objection. The first is to deny that those suffering from OCD experience their bodily movements as actions, which I have argued is a criterion for a sense of ownership to arise. Rather, these people may experience their bodily movements much like those suffering from Anarchic Hand Syndrome, and thus do not constitute an effective counterexample to the claim that the sense of control is necessary for the sense of ownership. I do not think that this is correct. The urges that an OCD sufferer experiences might be classified as unendorsed desires to perform certain bodily movements. For example, an urge to count to ten twice before taking a bite of that slice of pie can be described as really wanting to do so without wanting to want to do so. This, I think, is the sense of alienation that those who deny that OCD is accompanied by a
sense of ownership have in mind, but I do not think that this alienation is robust enough to warrant the claim that the bodily movements are not experienced as actions. As Anthony Marcel puts it: "the urge is felt as alien and coming from elsewhere [...] but the action itself is owned" (Marcel, 2003, p.80).

This brings us to the second way of handling the objection, i.e., that there is a sense of control over bodily movements in cases of OCD. The biggest piece of evidence for this comes from the fact that there are often very specific rituals that someone with OCD must perform in order to quell their urges. These rituals often consist of highly specific movements that must be initiated and terminated at very specific times, as well as carried out in a certain way. Of course, an individual suffering from OCD lacks a sense of control over the disorder itself, that is, she does not feel that she can control the occurrence of these compelling urges, but this is a different issue from whether or not she has control over the way in which she performs individual actions in response to these urges.\(^{12}\) Thus, cases of OCD should not prevent us from making the claim that the sense of control is necessary for the sense of ownership.

Having described the sense of control as the experience of being able to initiate or terminate an action, combined with the experience of being able to guide the action to a certain extent, I now turn to the question of what it takes in order for one to experience oneself as successfully guiding an action. A plausible assumption is that in order for one to have this experience, one must be aware of both one's intentions in performing those actions and one's bodily movements, so that one might keep track of how well one is executing the action one intends to perform. This, in turn, suggests that the sense of

\(^{12}\) Kirstie Laird pointed this out to me in a personal communication.
control over our actions is closely related to the awareness of our intentions and the awareness of our bodily movements. Here things start to get a little murky, however, for we can readily think of actions that are seemingly not preceded by conscious intentions and for which we still have a sense of control. Moreover, psychological research suggests that our awareness of our bodily movements is not as acute as we generally suppose. Thus, we must establish the role that the awareness of our intentions and bodily movements actually plays in bringing about a sense of control over an action.

How are we aware of our own bodily movements? Common sense dictates that we must at least have already made a move in order to know that we have done so. Surprisingly enough, however, according to the results of some studies in psychology, people may report being aware of performing a bodily movement before it has even occurred. For example, in the study conducted by Libet et al. (1983), when subjects were asked to judge the time at which they had performed an intentional hand movement (their \(M\) judgment) they had an average awareness of the action 86ms before its actual onset. Thus, it appears that subjects were not aware of the movements themselves, but, rather, their reports corresponded to pre-motor processing in the brains. Another study by Haggard (1999) further supports this conclusion. Subjects were required to type sequences of pre-rehearsed letters that were one, three, or five letters long, and were then asked to judge the time at which they pressed the first key in the sequence. The experimenters used the commonly accepted fact that longer sequences take more preparation, to test the hypothesis that the longer the sequence, the greater the error in the subjects’ judgments of their first key presses. Indeed, the results showed that the longer the key sequence, the more erroneously premature the perceived time of the first key
press. This suggests that the subjects’ reported awareness reflected the onset of the preparation for the action rather than the onset of the action itself.

It should be noted that there is still the mystery of exactly how these preparatory brain events and the conscious awareness that co-occurs with them are related. We would not want to say that subjects are aware of these brain events, since they need not even know that they have a brain in order to report this awareness.\textsuperscript{13} It might be safer to say that the brain events are neural correlates of the conscious experience of being aware of an action, or more strongly, that the conscious awareness is somehow reducible to the underlying brain events. The appropriate description however, is a broader metaphysical issue regarding the nature of the relationship between consciousness experience and the brain, and is, unfortunately, outside the scope of this project. The main point here is that reported awareness of the initiation of an action can occur prior to the action itself, and that this does not seem to affect any sense of control one might have over the action.

A further study described by Marcel (2003) suggests the sense of control can even be had when there is no reported awareness of a bodily movement. The study hinges on the fact that it is possible to induce flexion and extension of the arm by vibrating the biceps and triceps tendons at the elbow, respectively. Interestingly, if the induced movement is inhibited, the illusion that the elbow is in the position opposite to that it would occupy given the reflex is created, with the rest of the arm and hand following suit. Subjects are able to report the position that they think their arms are in by using a visual map or by using their other arm to demonstrate what they take the position of the affected

\textsuperscript{13} Marcel (2003) makes a similar point: “I prefer to refer to experience associated with such activity rather than experience of such activity, because I doubt that in such experience we are aware of the specifications or commands as such or under that description.” (Marcel, 2003, p.74)
arm to be. Most remarkable is that if a subject is asked to grasp the wrist of the stimulated arm with her other hand, she is able to do so correctly without realizing the disparity between the location she just grasped and the location in which she reported her stimulated arm to be.

In order to explore this phenomenon more thoroughly, Marcel (2003) had subjects put their elbows in a sling under an occluded surface. A number of lights that were arranged in the shape of an arc were then projected onto the surface, so as to suggest a possible path for the arm below the surface to take. The elbows of the subjects were stimulated so as to trigger the illusion described above, and one of the lights on the occluded surface would then go on. The subjects were asked to either draw the arm movement they would need to make to get their hand under the light and then move their hand under the light; move their hand under the light and then draw the movement they just made; or do all of the above. The results showed that in all cases, the pre-movement drawing followed the illusion, while the movement of the hand under the light was always immune to it, and by and large (in 60-70% of cases) the subjects’ pre-movement drawings matched their post-movement drawings. The other 30-40% of subjects made accurate post-movement drawings, many of them remarking that, “while they had intended to move in one direction, they had the impression that they had moved in the other” (Marcel, 2003, p.66). Since the individuals in this experiment reported nothing out of the ordinary about their sense of control over their actions, we can only assume that their experiences of agency were unaffected by the fact that the exact positions of their limbs was not consciously accessible to them.
What are we to make of all this? I suggested earlier that in order to have a sense of control over an action, it would be at least necessary to have an awareness of one’s bodily movement. It now seems possible that the awareness required might not depend on actual bodily movements, or even that conscious awareness might not be necessary after all, although given that the subjects in Marcel’s study were able to execute their intentions, it seems we should attribute some sort of awareness to them. I leave the discussion of bodily awareness here for now, as I have at least established that the awareness of the body involved in the experience of agency cannot be reduced to proprioception or visual or tactile observation, which is an interesting conclusion in itself and one that I can make use of in the following chapter.

I now turn to the claim that the awareness of an intention to act is necessary for a sense of control. Some deny that even this is so, since we can think of cases of ‘normal’ action in which we lack them altogether. Marcel, for example, claims that “[u]nawareness of intention appears to be more common than supposed. In so far as we are aware of our intention, it is hard to see how intention, in those cases where we are unaware of it, can be an ingredient of felt ownership of action or play a significant part in the sense of our causality of an action” (Marcel, 2003, p.61).

In order to support his claim that unawareness of intention is pervasive in everyday action, Marcel points out that we often “lack explicit awareness of sub-goals in achieving a goal” (Marcel, 2003, p.60). He gives as an example picking something up from a low table, where one may be aware of the goal, but unaware of whether he will do it by bending at the waist or the knee. This, to me, seems not to be an argument against the necessity of conscious intentions and more like an observation further bolstering the
conclusions just drawn about our somewhat weak awareness of exact bodily movements. Just like in the Marcel (2003) study, the subjects were not aware of the movements they needed to make in order to move their hand to the light, and they thus had no conscious intentions to make them, but they did have an explicit representation of a goal in acting. So it may very well be true that we are not always aware of the sub-goals involved in achieving a larger goal, but this is not inconsistent with the claim that in any situation in which we experience a sense of agency, we act with some goal in mind. The brain can leave the details of the action up to the motor system to resolve, but what is important is that we are somehow keeping track of the overarching goal with which we are currently occupied.

Of course, there are situations in which we patently lose track of our overarching goal and are only aware of a sub-goal. These cases signal “[a] loss of contact between the you that was in charge then and the you that is in charge now” (Dennett, 2003, p.253). For example, one might find oneself in the kitchen, knowing that one has gone there in order to do something, and yet have no idea what that something might be. These absentminded cases seem to support, rather than weaken, the claim that awareness of intentions is a necessary component of the experience of agency, for in such cases, when one finally comes to realize that one does not know what one is doing, one’s sense of agency is momentarily interrupted as one struggles to regain contact with one’s original intention. Other cases of absentmindedness may involve no prior intention at all. For example, one might suddenly find oneself tapping one’s foot to the rhythm of the song on the radio, although one had no intention whatsoever to do so. What should one say about these cases? It seems that one could reasonably say that these types of activities really
aren’t accompanied by a sense of agency, for although we may have access to them on some level, that is, if someone asks us to please stop tapping our foot, we can oblige the request, we might defend our behaviour by remarking that we “didn’t realize we were doing that” or something along those lines. Cases of absentmindedness, then, seem to support the claim that what is required for a robust sense of control is an awareness of one’s intention in acting.

The discussion above highlights the need for some clarification regarding the sense in which I am claiming that one is aware of one’s intentions. I do not wish to claim that prior to every single action, one must consciously rehearse one’s intention to act in order for us to be able to say that one is aware of one’s intention. In fact, I readily admit that it is rarely the case that we make explicit an intention to perform an action, except perhaps in situations where we are resolving to do something that we might otherwise be reluctant to do, for example when someone resolutely forms the intention to take the garbage out later. But this does not imply that one is not conscious of one’s intentions in all cases in which they are not made explicit in this manner. Rather, I propose that having a sense of control over an action requires only that one is aware, on some level, of what one plans to do. These plans may be quite general, and may leave the details of implementation largely unspecified, but they will always consist of an overarching goal that imposes satisfaction constraints on its execution. Moreover, the sub-goals involved in the goal execution may be carried out entirely by sub-personal processes, but the occurrence of these processes is entirely dependent on the larger goal in relation to which they are structured.
Consider again the intention to pick something up from a low table. It is most likely not important to one’s plans whether one does so by bending at the knees or the waist, so long as one is able to achieve the overall goal. As such, it is no surprise that the computational problem of which way to go about picking up the object might be solved by the motor system without the help of any conscious processes. If, however, there exists some sort of conflict between one of the possible sub-goals and another of the agent’s goals, then it is possible that the execution of the sub-goal would come under conscious awareness. For example, one might want to pick something up from a low table without drawing any attention to oneself, in which case the particular movements one engages in to achieve the overall goal will most likely be consciously planned.

One might object to the claim that we must have intentions to act even in this loose sense by appealing to a psychological phenomenon known as Utilization Behaviour, which most often affects those with a lesion in their frontal lobe. Individuals exhibiting Utilization Behaviour suffer from the inability to inhibit responses to environmental stimuli. Often, this behaviour is manifested in the usage of props for no reason other than that they are proximally present:

An examiner sits with the patient, for example, and at no time makes any suggestion or encouragement for the patient to act. Instead, the examiner touches the patient’s hands with a glass and carafe of water. Normal individuals in this situation typically sit there and do nothing, but frontal-damage patients may grasp the glass and pour it full from the carafe. Touching the patient’s hand with pack of cigarettes and a lighter prompts lighting a cigarette, and a variety of other objects may prompt their associated actions. One patient given three pairs of eyeglasses donned them in sequence and ended up wearing all three (Wegner, 2002, p.122).

The interesting thing to note here is that these individuals are fully aware of what they are doing and do not deny ownership of their actions. If they are asked for a reason for their
actions, they will either say that they have none, they will invent one, or they will say that they just “felt like doing it” (Marcel, 2003, p.77).

It seems that in these cases, there really can be no intention to act, since the ‘actions’ performed have the characteristics of environment-driven automatic responses. Just like we would not call the automatic constriction of the pupil in response to light intentional, we should not call the automatic use of a prop placed in one’s hand intentional. And yet, the sense of agency remains undisrupted despite this putative lack of intention to act. How then, can I make the claim that the sense of agency is inextricably linked to the awareness of intention as I have above?

It is difficult, if not impossible, to draw concrete conclusions regarding the phenomenological experiences of individuals exhibiting Utilization Behaviour, but I would like to suggest some interpretations of their behaviour and their verbal reports that are not inconsistent with the claim that if they have an undisrupted sense of agency, then they must be acting intentionally. First, I should point out that just because these individuals often deny having a reason to act, this does not imply that they do not have an intention to act. I might jump and click my heels in the air while walking down the street quite intentionally but for no particular reason, and in this situation, as in many others, it would be perfectly appropriate to describe my behaviour as something I just “felt like doing”. Likewise, it might be the case that a patient that dons three pairs of eyeglasses in sequence intends to do so, even if there is no good reason. Here one might protest and remind us of the fact that these seem more like automatic responses rather than intended action. In response to this, we might draw parallels with the actions of those with OCD. Indeed, these individuals cannot inhibit the urge to act in a certain way, but once that urge
is there, they form a perfectly valid intention to act in accordance with it. Similarly, those with Utilization Behaviour might harbour urges to respond to objects in their environment and form intentions to do so on the basis of those urges.

Second, it is possible that those with Utilization Behaviour are not able to express any reaction to their behaviour like one might expect\textsuperscript{14}, not because there is nothing qualitatively different in their experience of performing them or because they do not recognize their urges as bizarre, but because they are unable to show any affect at all (Marcel, 2003, p.77). Indeed, an all around “lack of concern” is a typical disposition of those with Utilization Behaviour (Marcel, 2003, p.77), so it might very well be the case that it extends to their uninhibited use of props. Of course, none of this speculation is conclusive, but I think that I have at least made room for the supposition that those with Utilization Behaviour either do not experience a ‘normal’ sense of agency or do indeed have intentions to act in the way that they do.

In this chapter I have attempted to offer a clearer picture of the content of the experience of agency. I have argued that it consists mainly of four interrelated components: (i) the sense of ownership, (ii) the sense of control, (iii) an awareness of intention, and (iv) an awareness of bodily movement. I have also argued that the sense of control is derived from an awareness of intention and a qualified sense of bodily awareness, and that this in turn is necessary for a sense of ownership to arise. I now turn to the question of what the experience of agency, described as such, can tell us about our selves.

\textsuperscript{14} Anarchic Hand patients, for example, often express terror at the unwanted movements of their afflicted hands (Marcel, 2003, p.77).
THE EXPERIENCE OF AGENCY AND THE SELF

Having offered a description of the content of the experience, we are now in a better position to explore it from what I think will prove to be a fruitful perspective: the role it plays in the understanding of the self. More specifically, the claim I wish to consider in this section is that the experience of agency constitutes a form of self-knowledge. I should make it clear that I am not here concerned with defending the possibility of a conscious experience being an epistemic source in the first place, but rather I assume, quite reasonably I think, that it can be and I turn my attention to the claim that the knowledge we derive from the experience of agency is self-knowledge.

Before considering how it is that the experience of agency can provide us with self-knowledge, it will be useful to observe that self-knowledge and agency in general, that is, without reference to the experience thereof, are themselves related. After all, it is not possible for one to act intentionally without knowing that one is doing so, and this knowledge must at least consist in knowing what one is doing, which is a piece of knowledge about the self. Donald Davidson (1971) recognizes this when he says that

[a]ction does require that what the agent does is intentional under some description, and this in turn requires, I think, that what the agent does is known to him under some description. But this condition is met by our examples. A man who raises his arm both intends to do with his body whatever is needed to make his arm go up and knows that he is doing so. And of course the cerebral events and movements of the muscles are just what is needed. So, though the agent may not know the names or

15 My use of the term 'self' here should only be taken to refer to our self-concept, viz. that way in which we conceive of ourselves, and I am not thereby endeavouring to make any metaphysical claims about the nature of the self. Whether or not our self-concept ends up exhausting the metaphysics of the self is an issue that I put aside for others to settle.
locations of the relevant muscles, nor even know he has a brain, what he makes happen in his brain and muscles when he moves his arm is, under one description, something he intends and knows about (Davidson, 1971, p.50, my emphasis).

I agree with Davidson that a certain degree of self-knowledge is necessary for one to act intentionally, namely in the form of knowing what you are currently doing, but a more controversial claim, and something that Davidson does not venture to say, is that the way that an agent knows about her actions is via the experience of agency.

Roughly put, the experience of agency provides us with the knowledge that we are acting. This is evidently a form of self-knowledge, but knowledge in general can also be categorized in terms of its sources, e.g. observational vs. non-observational, and the form it takes, e.g. propositional vs. procedural. The knowledge that I am acting is a species of non-observational, propositional knowledge. It is non-observational because it is not acquired from a third-person perspective, but rather a first-person experience. It has been described by some as ‘knowledge from the inside’ (see, for example, Campbell, 2003; O’Brien, 2003; Eilan & Roessler, 2003), further highlighting its fundamentally subjective character. As has been demonstrated by the cases of Anarchic Hand Syndrome, it is impossible for one to know for certain, from observation alone, that someone else is acting, i.e. not just performing bodily movements but full-fledged intentional actions. Granted, we can be fairly good at judging when it is the case that someone is performing an action by simply observing them, but ultimately the defining feature of the action is the qualitative experience with which it is accompanied and to which only the person performing the action is privy.
The knowledge that I am acting is propositional, for in its broadest sense, its content is the proposition *that* I am acting. But this is not all we know through the experience of agency, for it also informs us of what we are doing in relation to our intentions. For example, when I reach out for my teacup intentionally I am imbued with knowledge of my *intention* to do so and the fact that I *am* doing so, not just the knowledge that I am doing *something*. Thus, we must filter the broader knowledge claim that, via the experience of agency, I know that I am acting, into the two narrower claims that I know what I intend to do and I know what I am doing.

Here a special worry might surface, for I argued in the last section that the experience of agency can sometimes be generated in situations in which the agent seems not to be able to tell the actual position of her body, as demonstrated by Marcel (2003). How, then, can I make the claim that the subjective feeling accompanying our actions provides us with knowledge of what we are doing? In order to address this concern, I must qualify the claim that knowledge of what we are doing amounts to propositional knowledge. *In some cases* the knowledge of what I am doing may manifest itself in a way that is not propositional. In these special cases, the determining factor for whether I actually know what I am doing is not that I can express a belief in the proposition corresponding to my action, but rather that I do what I intend to do. Thus, in the case of the Marcel (2003) study in which the subjects were unable to accurately express the actual position of their forearms and yet were able to successfully perform an action in accordance with their intention, we would, and I think should, say that they have knowledge of what they are doing.
How exactly is the knowledge of what I intend to do and of what I am doing related to the self? The first thing to note is that there are several types of facts that I can know about myself. For example, I can know autobiographical facts, facts about my behavioural dispositions, facts about my physical traits, facts about my social circle, facts about my likes and dislikes, etc. These categories of facts can be seen to correspond to certain modules of the more general self. For example, autobiographical facts, which include memories, experiences, etc., correspond to one particular component of the overall self that can be labelled the 'autobiographical self'. This sub-self is inseparable from the general self, but is nevertheless specialized enough that it can be conceptualized on its own.

Here one might ask to which modular self does knowledge of what I am doing and of what I intend to do apply. I argue that knowledge of what I am doing is knowledge of the embodied self, and knowledge of what I intend to do is knowledge of the rational self. By 'embodied self', I mean the identification with one's physical body through a representation of it. This does not merely involve conceiving oneself as having a physical body, but rather conceiving oneself as a physical body.

How do we represent our physical bodies to ourselves? The neuroscientist Antonio Damasio (1999) suggests that this feat is accomplished by a set of specialized areas in the brain designed to map the state of the body. Damasio distinguishes between two manifestations of the embodied self: the proto self and the core self. The proto self and the core self work in tandem to constantly monitor the dynamic state of the body,

---

16 The idea of modular selves is not new; indeed, the psychologist William James posited four different constituent selves of the entity he described as 'me'.
with the proto self mapping the body state directly by representing it via neural patterns\textsuperscript{17}, and the core self representing, on a conscious yet non-verbal level, any change in the proto self that occurs through bodily movement. In this way, we are aware of the state of our bodies and can make use of this information to successfully execute our actions. Moreover, the core self gives rise to what Damasio describes as a “sense of self about one moment – now – and about one place – here” (Damasio, 1999, p.15).

As is so often the case, we can gain a clearer understanding of the normal functioning of this sense of self by considering scenarios in which it is impaired, for example in the case of patients with what is known as ‘complete anosognosia’. Complete anosognosia is a neuropsychological disorder that manifests itself in the inability to recognize one’s physical ailments. Damasio describes a particularly striking case as follows:

Whenever I asked my patient DJ about her left-side paralysis, which was complete, she would always begin by saying that her movements were entirely normal, that perhaps they had once been impaired but they no longer were. When I would ask her to move her left arm, she would search around for it and, after looking at the inert left limb, ask whether I really wanted “it” to move “by itself.” When I would say yes, please, she would then take visual notice of the lack of any motion in the arm, and tell me that “it doesn’t seem to do much by itself.” As a sign of cooperation, she would offer to have the good hand move the bad arm: “I can move it with my right hand.” (Damasio, 1994, p.62).

Damasio argues that this deficient sense of self is a by-product of the “impaired ability to process current body states” (Damasio, 1994, p.237), which, in turn, results from damage to the neural substrates underlying the proto self and the core self (Damasio, 1994).

\textsuperscript{17} I leave the question of how it is that neural patterns can represent for another project.
Damasio’s (1999) account of the embodied self involves a conscious representation of all the minute changes in one’s overall body state, of which the sense of self is the phenomenological component. However, it is not the case that we experience ourselves as the agents of all the bodily changes of which we are capable of being aware. Rather, the experience of agency only constitutes epistemic access to a certain subset of those changes, namely those that make up our actions. It is this knowledge of our bodily movements as actions that makes the experience of agency a special variation of the knowledge we have of all our bodily changes via the proto self and the core self, but nonetheless links it epistemically to the embodied self.

How is it that we are able to have, via the experience of agency, selective knowledge of our actions that is irreducible to the knowledge that we have of our bodily movements in general? This is where the knowledge of what I intend to do, which amounts to knowledge of what I have labelled the ‘rational self’, comes in. The rational self, as I am using the term, is anchored in the representation of an agent’s intentions and identification with those intentions. Simply put, it is the part of the self that we conceive of as involving our intentions, our future plans, and our reasons for performing actions. The experience of agency partly constitutes knowledge of the activities of the rational self, and it is this knowledge that extends into the selective knowledge of our actions. In other words, by knowing the rational self, we come to know the embodied self, and it is by way of the experience of agency that the two are linked.

This unique self-knowledge that we get from the experience of agency also plays a role in informing our most basic sense of self, by facilitating the division between the self and the mind-independent external world (Russell, 1995). The behaviour of infants
under certain controlled conditions suggests that they take their subjective experience to be not just one viewpoint among many possible perspectives of an objective world that exists independently of them, but rather an exhaustive view of reality, i.e. when their head turns, the world turns, when an object exits their visual field, it ceases to exist, etc. The experimental psychologist James Russell (1995) argues that an experience of agency is instrumental in the birth of our self-concept by prompting the recognition of the distinction between the subject and the mind-independent external world in infancy, and thus delineating the boundaries within which the self exists.

This argument hinges on the very simple and uncontroversial observation that we all have a sense of agency for some events in the world but not for others. For example, I have a sense of agency for the movements of my right hand as I strum my guitar and my left hand as it makes its way across the fret board, but in contrast, I do not have such an experience for the movement of the sun across the sky, or for the leaf spiralling to the ground. It is this distinction between experiencing a sense of agency and lacking one altogether that grounds the parallel distinction between the self and the world for an infant. More specifically, the sense of agency aids in the development of the distinction because it allows infants to: (i) make preliminary strides in distinguishing between how the world appears and how it is in fact through bodily awareness, (ii) recognize the special mode of non-observational knowledge they have of their own actions versus that they have of events in the external world, and (iii) recognize the reversibility of the production of their own actions versus the irreversibility of events in the external world (Russell, 1995).
In order for infants to be able to process perceptual information in a way that allows them to recognize that there is a difference between the way the world appears to them and the way that it is in actuality, they must be able to distinguish between situations in which an object is actually moving across the visual field and situations in which there has merely been a shift in visual attention or head orientation (Russell, 1995). This turns out to be a somewhat difficult computational problem for the visual system to solve, but what is crucial to its solution is that the infant be aware of its own bodily movements, so that it may calculate the difference between its own movements and those independent of them. The sense of agency, which provides humans with knowledge 'from the inside' of the activities of the body, is thus paramount to the development of the ability to tell the difference between the movement of the objects in the world and one's own movements.

The contrast between this knowledge 'from the inside' and the third-person, observational knowledge that we usually enjoy further serves to sharpen the distinction between self and world. Infants recognize that they know about their own actions in a completely different way than that in which they know about the activities of the household pet, or the toy truck, and that the difference lies in the fact that in the former case, they experience having direct epistemic access to their intentions and actions, and in the second case they experience having to rely on observation.

Events in the world occur in a certain order, regardless of whether or not we are there to perceive them. If a ball rolls by and then a child chases it out on to the street and almost gets hit by a car, there is no way to reverse this sequence of events. Through the special kind of control I experience myself having over my actions as an agent, however,
I experience a certain degree of freedom with regards to the order in which I receive perceptual inputs. For example, when looking at a painting, I can choose to look at any part of it in any order I please. There is nothing stopping me from examining it from top to bottom, from left to right, or in a more chaotic fashion. The order in which I explore my visual field is not fixed like the order of events in the world, and the sense of control I have over my exploration of my visual field is what highlights this special fact (Russell, 1995).

These three things – bodily awareness, non-observational knowledge of action, and a sense of control over the order of perception – are all components of the overall experience of agency infants have for their actions, and assist in the recognition of the division between the self and the world. This separation is the first step in acquiring one’s own self-concept. Separating self from world, however, is just one of the first steps in the development of the robust and complex self-concept that most human adults come to possess. The experience of agency contributes to the development of two of the most important and complex self modules that we come to have: the ‘social self’ and the ‘moral self’.

I am taking the concept of the social self to revolve around our ability to communicate and co-operate with other agents. For the most part, co-operation hinges on the ability to communicate what it is that we plan on doing and what it is that we want others to do, and this, in turn, requires knowledge of our intentions and our actions of the sort that the experience of agency provides most efficiently. Picture getting through daily tasks without this knowledge: there would be no future plans; no making promises; no asking for help; and essentially no co-ordination between people. The capacity to
monitor our plans and actions is an extremely fitness-enhancing trait, allowing us to co-operate with our conspecifics on an entirely new level. As Dennett puts it,

[w]e have added a layer on top of the bird’s (and the ape’s and the dolphin’s) capacity to decide what to do next. It is not an anatomical layer in the brain, but a functional layer, a virtual layer composed somehow in the micro-details of the brain’s anatomy: We can ask each other to do things, and we can ask ourselves to do things. And at least sometimes we readily comply with these requests [...] We human beings not only can do things when requested to do them; we can answer inquiries about what we are doing and why. We can engage in the practice of asking, and giving, reasons. It is this kind of asking, which we can also direct to ourselves, that creates the special category of voluntary actions that sets us apart (Dennett, 2003, p.251)

The development of this social self by which we may forge co-operative relationships with one another and follow certain codes of behaviour, moral or otherwise, within communities of agents, would not be possible were it nor for the experience of agency and the knowledge with which it provides us. One might here wonder why it need feel like anything in order for us to recognize our intentions and our actions as our own. Here Wegner offers a felicitous response, despite some of the unwarranted conclusions he draws en route to it:

Conscious will is the somatic marker of personal authorship, an emotion that authenticates the action’s owner as the self [...] [t]his helps us tell the difference between things we’re doing and all the other things that are happening in and around us. In the melee of actions that occur in daily life, and in the social interaction of self with others, this body-based signature is a highly useful tool. We resonate with what we do, whereas we only notice what otherwise happens or what others have done. Thus, we can keep track of our own contributions without pencils or tally sheets” (Wegner, 2002, p.327).

Once we can keep track of our actions and communicate our intentions to other agents, in effect engendering the social self, we conceive of ourselves as moral agents as well. Features of the moral self-concept are not restricted to the performance of actions
within a social context, but also include a sense of moral responsibility for what we do. Now the questions of what one is doing and why take on a new flavour, as they are used to inform our ascriptions of moral responsibility rather than just to facilitate co-operation between agents. Our own actions become candidates for reactive attitudes of praise and blame and subjective feelings of guilt and pride. But on what basis do we take ourselves to be morally responsible agents rather than just agents? It seems that the answer to this question lies in the sense of control, a component of the experience of agency, that we have over our goals and our actions. After all, we are not merely passive bystanders to the various goals we entertain and the actions that may result from attempts to achieve them; rather we experience ourselves as playing an active role in the formulation of those goals and in selecting between alternative courses of action that will lead to their ultimate implementation. But this is all just a way of saying that we experience them as being under our control, at least within reasonable limits, i.e. there may be external factors over which we have no control that constrain our actions and goals. This sense of control over our actions is central to the notion that we can always choose to do otherwise than we ultimately choose to do, which in turn underlies many conceptions of freedom and moral responsibility, and thus contributes to our sense of ourselves as moral agents.

In this chapter I have tried to show how the experience of agency can be viewed as a valuable source of self-knowledge. Not only does it offer us knowledge of our intentions and our actions, thereby acquainting us with our rational and embodied selves, but it aids in the very creation of a self-concept and in the development of this self-concept into more complex forms, such as the social self and the moral self.
CONCLUSION

My aim in this thesis has been threefold: (i) to evaluate the main stances on the veridicality of the experience of agency, (ii) to sketch an account of the rich phenomenological content of the experience itself, and (iii) to take some steps in a new direction that takes as its focal point the idea that the experience of agency is a source of self-knowledge that is fundamental to our self-concept.

First, in order to bring into sharper focus the contours of the veridicality debate, I dealt with proponents and opponents of the view that the experience of agency accurately represents us as the causes of our actions. On the proponent side, I considered the work of agent-causationists, who attempt to vindicate the experience of agency by translating it directly into a naturalistically plausible metaphysics that takes as its central tenet the notion that an agent, rather than an event, can be the cause of an action. I argued that a promising way to get around the metaphysical obstacles that an agent-causationist might face is to ground agent-causation within an emergentist framework, like the philosopher Timothy O'Connor does. I further argued that there is reason to doubt O'Connor's account, since it appears to be tailor-made for explaining our abilities to cause actions as agents in the very way that our experience of agency suggests that we do. I concluded that agent-causation is left on shaky ground and that agent-causationists take their commitment to the veridicality of the experience of agency too far.

On the opponent side, I dealt with the attempts of Daniel Wegner and Benjamin Libet to expose the experience of agency as an illusion. I suggested that Wegner needs to
offer a concrete example of an illusion of control in order to defend his thesis that the
experience of agency is an illusion because it misrepresents our intentions as the causes
of our actions. I further argued that he fails to offer such an example, and that even if he
did, it would merely show that our experience of agency, just like our visual experience,
can sometimes, not always, be mistaken. I then looked critically at Libet’s study from
which he draws the conclusion that our experience of agency is an illusion since we do
not actually initiate our actions, but rather they are initiated by events in the brain
preceding our conscious intentions to act. I suggested three possible interpretations of the
results of the experiment, settling finally on the Autopilot Interpretation, that suggests
that the subjects’ formation of a general prior intention to act in accordance with the
instructions of the study, and the simplicity of the motor act required of subjects should
be sufficient to quell our worries about the subjects’ agency being thwarted and the
experience of agency being an illusion. I concluded that neither Wegner nor Libet are
successful in their attempts to demonstrate that the experience of agency is an illusion.

Next, I undertook a shift in focus from the veridicality of the experience of
agency, to the phenomenon itself and its role as a source of self-knowledge. I then
proceeded to sketch an account of the content of the experience of agency, identifying
four related components: (i) a sense of ownership, (ii) a sense of control, (iii) an
awareness of intention, and (iv) an awareness of bodily movement. I argued that the
awareness of intention and awareness of bodily movement in a qualified sense are both
necessary for a sense of control to arise, and that the sense of control is, in turn, necessary
for a sense of ownership.
Finally, I turned to the question of how the experience of agency, described as having these four main components, could be viewed as a source of self-knowledge. I argued that it provides us with direct knowledge 'from the inside' of the activities of the rational self and the embodied self. I argued further that it is integral to the grounding of our self-concept by facilitating the recognition of the division between self and world in the early stages of infancy. I then argued that it continues to inform our self-concept by aiding in the development of the social self and the moral self. There is much more to be said about the experience of agency as a source of self-knowledge, but having laid down some foundations for future research, I leave things here.
REFERENCE LIST


75


Libet, B. (1999). Do we have free will? In B. Libet, A. Freeman, & J.K.B. Sutherland (Eds.), The volitional brain: Towards a neuroscience of free will (p.47-58). Exeter, UK: Imprint Academic.


