

Consciousness and the Will

By

Donovan Hulse

A Thesis submitted to the Faculty of
Graduate Studies in Partial Fulfillment
of the Requirements of the Degree of

Masters of Arts

Department of Philosophy
University of Manitoba
Winnipeg, Manitoba

© Donovan Hulse, August, 2004

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

Consciousness and the Will

BY

Donovan Hulse

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of
MASTER OF ARTS**

Donovan Hulse © 2004

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Consciousness and the Will

Chapter 1

In this thesis I will defend the position that consciousness does not have the causal efficacy over the will that so many people think that it does. This position will be motivated by both empirical evidence and argument. What I hope to show, in the end, is that consciousness is only related to the will in a very specific way. It will turn out, if I am right, that the role consciousness plays in willful action isn't anything like the role most people take it to play. This is because those things which we normally construe as conscious willful decisions are in fact being made by a non-conscious part of our brains.

No doubt my position will seem counterintuitive to many. It flies in the face of our everyday phenomenology. It certainly feels as though we are consciously directing our actions. But I hope to be able to account for the phenomenology in a way that will make it seem plausible that our phenomenology is tracking the right thing and that, even so, consciousness does not directly control action.

A bit of a caveat first. I am an agnostic as to whether the will is determined or not, so in this thesis I will neither defend nor attack the view that our will is free. All of my arguments will center on whether or not the will is a function of consciousness or not. None of this will bear on whether the will is causally determined according to the laws of physics or if we in fact can make free choices.

The layout for this thesis will be the following: First I will introduce what I take to be everyday characterizations of consciousness and the will. I will then present a puzzle about the will that should stir one's intuitions about its relation to consciousness. In the following chapter I will present arguments from other philosophers that claim that the will and consciousness must be related, and another argument that they must not be. Then, in the next chapter, I will present empirical evidence meant to demonstrate how, in our brains, consciousness is related to the will. In the next chapter I will offer an argument that will further motivate the position that the will and consciousness cannot be related. In the end all of these arguments and pieces of evidence will be brought together to provide a coherent, justified theory that should trump those who argue the will and consciousness must be related, if only because I will offer what I take to be the best explanation of the phenomenology of willful actions, the empirical evidence and arguments.

The Will

When we talk about the will, we usually mean that part of us that gets us to act. One often needs to will oneself out of bed in the morning. We wake up groggy and finally come to realize where we are. When we finally make up our minds about what we are going to do we have a feeling of exerting ourselves to get our bodies out of bed. It is this feeling of pushing ourselves to do something that people usually associate with exerting their wills.

Most of our actions in day to day life are considered to be produced by the will. This is because the will is taken by many to be the seat of action. Without a will our

actions wouldn't be our actions, they would be the product of some process over which we have no control.

It is also common to talk about people as having more or less willpower than other people. It seems to me that this must be a way of talking about stronger or weaker desires that a person has. Certainly there are situations, like getting out of bed, where I feel that I am exerting my will more than I am in other situations. Everyone has things they must force themselves to do and that feeling of taking control and forcing oneself to act in a certain, often undesirable way, is often considered to be a function of the will.

At the theoretical level the will is usually associated with our desires and intentions. However there are a number of other mental states that could be seen as causing us to act. Wishes might be something that motivates a person, as might goals or tryings or any number of other mental states, usually referred to as "pro attitudes". Any one of these mental states might initiate action in us and so might be construed as part of the will. Desires are the paradigmatic pro attitude as they are most clearly associated with action production.

There is some confusion about how, exactly, to characterize the will. Is it the desire that causes us to act or do desires need to be supplemented with intentions in order to have casual efficacy? One reason to think that desires might require intentions before action can occur is that we seem to be able to have desires for a very long time before

ever acting on them. One's desire to go to grad school can be present, but not directly acted on, throughout one's undergraduate degree. But it isn't clear that intentions solve this problem, because if intentions are needed for action then they seem to need a desire to produce them. So intentions may be what actually cause the actions but intentions may not exist if not caused by a particular desire.

In what follows I will be characterizing the will as a collection of desires and intentions. But not all of our desires count as part of the will. If I have a desire, say to get a job at MIT in 4 years, then only if I come to act on that desire does it count as part of the will. As noted above, only those mental states that cause us to act count as part of the will. If mental states that don't cause us to act can be seen as part of the will then it seems possible that I could will something to happen without actually doing anything about it, an odd consequence indeed in someone who isn't paralyzed. Furthermore, not all action-producing mental states count as part of the will since reflexive actions don't count as willful actions. So perhaps it is best to say that the will is that collection of mental states that causes us to act intentionally. But what, exactly, those mental states those are is an issue still up for debate.

So where to actually put the sign post that indicates where the will is, I am not sure. Perhaps at desires, perhaps at intentions or perhaps at some larger collection of our pro attitudes. Since there is no definitive way to demonstrate which of these mental states is to be associated with the will I will remain theoretically neutral on this point.

Lastly, some might think that the will is the place in us where decisions on how to act are made. There is a sense in which I take this to be true and a different sense in which I take it to be false. The sense in which I take it to be false is the sense in which, say, a person is sitting around trying to decide what movie to see and she scans the paper, reads some reviews, weighs her preferences and then come to a definitive decision about which movie to go to. As noted above, I take none of this to be a product of the will because none of it has done anything to actually get the person off of the couch to go see a movie, insofar as decision making doesn't entail anything about acting on that decision. Since the will, I take it, is what gets us to act intentionally and no intentional action has been taken (except reading) I don't think anything willful has been done, as far as movie going is concerned. One might find this puzzling: after all the decision has been made, and what could be more willful than that? The problem is that for any decision of this sort it could just as easily be acted on as not. If, in a moment of bad decision making, I decide to buy some heroin when I get paid on Friday, does that count as a willful decision? I hope not as I will more than likely never act on it and if it does count as willful then it seems like the following could be true: I could will myself to do something (that isn't impeded in any way) that I never do. I take this to be contrary to the will's role as the seat of intentional action. So all of these conscious decisions that we make in day to day life that we may or may not act on don't count as instances of willing, just because they don't involve intentions to act in any way. They involve decision making and decision making doesn't entail action. Basically, I am making the claim here that

there is an important distinction between decisions and motivation and decisions need not motivate us but that the will does.

The sense in which I think the above assertion about decision and the will is right is that our brain can act in any number of ways to satisfy our desires and intentions. So part of what the will has to be able to do is select the best action and form an intention to act that way. But, as we will see in Chapter 3, none of this decision making involves consciousness in the way that coming to a decision about movie attendance does.

Consciousness

In normal parlance people don't usually speak of consciousness but rather of things of which they are conscious. People talk about being conscious of a painting, of how much their feet hurt, the sound of a Bach concerto or of a memory of a past experience. All of these experiences are part of consciousness. Our consciousness is made of just these kinds of things, experiences. When people talk of being conscious of something they mean that there is a way it feels for them to be having that particular experience. In some literal sense, I cannot have your experiences and you cannot have mine. This isn't to say that two people cannot experience the same thing. Two people can look at the same painting from the same angle and have the light hit their retinas in the same way. The experience itself, the way it feels to the person

looking at the painting, will be different for each person but only qualitatively because the object being experienced will be the same in both cases.

In the philosophical literature these subjective qualities of experience are usually referred to as “qualia”. It is these qualia that philosophers usually talk about when they talk about consciousness. This is because it is a complete mystery how something physical, like your brain, could give rise to something subjective like qualia, and yet consciousness seems to be the product of such a process. Further characterizations (by David Chalmers) of the subjectivity of experience include the following:

A number of alternative terms and phrases pick out approximately the same class of phenomena as ‘consciousness’ in its central sense. These include ‘experience’, ‘qualia’ ‘phenomenology’, ‘phenomenal’, ‘subjective experience’ and ‘what it is like’. Apart from grammatical differences, the difference among these terms are mostly subtle matters of connotation. ‘To be conscious’ in this sense is roughly synonymous with ‘to have qualia,’ ‘to have subjective experience’ and so on.
(Chalmers, 1998, 5)

So when people talk about being conscious or having a conscious experience they are talking about having some kind of subjective experience. Hopefully this captures

what most people are talking about when they talk about their conscious lives. When I talk about consciousness in what follows, I will be talking about the same thing.

Consciousness and the Will

Since I am trying to describe the will and consciousness in common parlance it is important to add that sometimes terms like “conscious action” and “willful action” are treated as if synonymous. That is, people treat consciousness and the will as if the two were, in some ways, the same thing. It seems common to think that if someone was consciously trying to open a window then that action must have been a product of her will (And vice versa, that if someone willfully committed an act then he must have done so consciously not merely have been a conscious observer of it). Obviously, I take common parlance to be dead wrong about this, but it is important to characterize just the kind of intuition I will be arguing against in what follows.

Of course having said all of this it is not as if consciousness bears no relation to the will. Consciousness, on my account, feeds information to the will so that the will can decide how to act. But the relationship between the will and consciousness isn’t one of identity (the will and consciousness are not the same thing). The relationship is, loosely, more like the relationship between the hard drive and processor in your computer, information is swapped back and forth but each has a distinct role to play. The role of consciousness is to tell us what is going on. The role of the will is to initiate intentional action. But like the hard drive and processor, we shouldn’t take the

will and consciousness to be the same thing simply because they share information with each other, as each has a distinct role to play.

*Libet's Puzzle*¹

Now that I have laid the groundwork for what I'll be talking about in what follows I want to present a puzzle about the will and consciousness. I hope that this puzzle will stir people's intuitions about how the will and consciousness are related and motivate people to think about whether or not the two are separable.

Imagine a student sitting in class raising her hand to ask a question. If you asked her whether she is actually in conscious control of her hand raising, she would almost definitely say, "Yes!" The position I take to be indicated by Libet's results is that the correct answer would be "No!" We are not, in the most natural sense of the expression, in conscious control of our actions. How could this possibly be? How could it be that we so regularly feel as though we are consciously controlling our actions? Libet's experiment into the relationship between our conscious feelings of volition and how we produce action will answer these questions to a certain extent. Hopefully some philosophical clarification on my part will make Libet's findings

¹ It should be noted that the experiment I am about to make reference to has never been replicated, involved a small sample size and to this day its methods and conclusions are much debated among both philosophers and neurophysiologists.

seem even more plausible. First though, we should look Libet's experiment and see what it has to tell us².

The experiment began with five subjects, each of them undergoing 40 trials of the experiment. These subjects were connected to an EEG, a device that measures brain activity through electrical changes at the scalp. The subjects were then asked to flex a wrist whenever they felt like it. The EEG was used to record brain activity in the brain before and during the flexion of the wrist. This allowed the timing of these readiness potentials³ to be compared to the subject's conscious intentions to move.

The subjects were also asked to indicate when they formed the volition to move. The experimenters placed a dial with a revolving light that rotated every 2.53 seconds in front of the subjects. The subjects would report where the light was on the rotating dial when they formed the conscious volition to act. Although this may seem like an imprecise way to time conscious volitions, it is important that things were done in this manner. If the subjects had been asked to push a button when they became conscious of the volition to act, this would have interfered with the experiment, since the subjects would have had more than one volition: the volition to push the button and the volition to flex the wrist. In addition, because the goal of the experiment was to test the relation between consciousness and volition, it would have been suspect to

² Libet frequently talks of non-conscious and conscious intentions. I have chosen to talk instead about volitions. I am neutral as to whether *all* intentions are part of the will but am sure that, by definition, all volitions are. So talking about volitions will make this discussion more applicable to the will than it otherwise might be.

³ Readiness potentials (frequently abbreviated as RP) are increasing levels of electrical activity in certain parts of the brain. The presence of these RP's is taken as indicative of brain activity in that area.

choose some method other than first-person reports of conscious volition to measure this relation.

The actual flexing of the wrist was also measured as a way to provide a stop point for the chain of events culminating in the action. "The subjects were free, however, to choose to perform this act at any time the desire, urge, decision and will should arise in them." (Libet, 1985, p. 530) The experimenters had all of the information necessary to judge the relationship between consciousness and volition. They had the time of the activity in the brain revealed by the EEG, the time of the experience of consciously intending to move and the measurement of the muscle activity in the wrist to indicate the carrying out of the volition. Everything was set to discover whether our conscious volitions occur before or after action-controlling brain activity.

The results of the experiments are as follows:

The RP onset time was found to be consistently in advance of W, the time of initial awareness of wanting to move. For all series in which all 40 acts were experienced as fully spontaneous and unplanned, the average RP [Readiness Potential] onset was about -535ms relative to the initiation of muscle action (as indicated by EMG). Reported times of conscious intention to act (W) in these same series with type II RPs [those not involving pre-planning] averaged about -190ms. The average onset of these RPs therefore *precedes* average W by about 345ms. (Libet 1985, 533)

This timing relationship, with the “physical” (cerebral processes) preceding the “mental” (conscious intention), held not just for all series but for each individual series of 40 self-initiated acts in which RP and W were recorded simultaneously. Although RPs of 40 events were averaged to produce the recorded RP, statistical and mathematical evaluation of the experimental data strongly supported the view that each individual RP precedes each conscious urge. (Libet 1985, 534)

All of this is a bit technical. What the experimenters found is that our brain activity starts in the motor area about 535ms before the action occurs, and that our becoming conscious of our volition to act occurs about 190ms before we act. But this means, as noted above, that there is a gap of about 345ms between our brain activity that initiates action and our conscious volitions and that *the brain activity comes before the forming of any conscious volition*. To put this in chronological order, first comes the activity in the motor area, then comes our conscious experience of the volition to act, and then we act. Since the will is the seat of volitional action and consciousness seems to lag behind the forming of the actual volition to act one might take this as indicative of the separation of the will from consciousness.

It is also important to note that these results generalize to all trials and all subjects. It isn't the case that sometimes the readiness potential came before the conscious volition and sometimes it did not. In all trials, the conscious volition was preceded by the readiness potential. It is true that timing did differ between the onset of the

readiness potential and the conscious volition. This can be attributed to the fact that there is a certain amount of error implicit in all uses of first-person reporting in scientific experiments.

What this experiment suggests is that a non-conscious area of your brain initiates action and not consciousness. The non-conscious readiness potential, not your conscious volition, causes you to act. This is why so many people consider the experiment revolutionary. For the first time there is evidence that we, our conscious selves, are not in control of our actions. This is the puzzle that Libet's experiment leaves us with: our phenomenology tells us one thing, that we are in conscious control of our actions, and the experimental evidence tells us another, that we do not consciously control our actions. What I will do is defend this result from a number of philosophical objections made against it. This is meant to clarify the position I am forwarding as well as to make the position seem more plausible. More will be said about this in the final chapter where I will go to greater lengths to defend Libet's experiment. Again, what follows is meant to clarify, not to provide a robust defense of Libet's work.

A number of commentators⁴ (Wood, Jung, Jasper and MacKay) hold the following view: The way to avoid the implications of Libet's experiment is to identify the mental state, the conscious intention, with the readiness potential. This would alleviate the worry that conscious intentions have no causal efficacy. If the readiness

⁴ Libet's article was originally published in *Behavioral and Brain Sciences* which includes commentary from people in related fields immediately preceding its published articles.

potential just is the neural substrate of the conscious intention, then the intention has all the causal power of the readiness potential. So the conscious intention retains all of the causal power normally associated with it. That would be enough to retain our normal view of how consciousness initiates action. As Wood puts it:

In contrast, if we assume that conscious intention is one of the many brain processes that contribute to the initiation of voluntary movement, then the explanation for the obtained result is straightforward. According to this view, the brain process(es) that mediate the conscious intention to act *must* begin before subjects can report that they are aware of that intention. To assume otherwise is to assume that conscious intention arises full-blown, out of nothing, instantaneously (“the Devil made me do it”), a prospect that is decidedly dualistic. Unless conscious experience is totally unlike every other physical process we know anything about, it must have a non-zero time course; if its time course is anything like that of other brain processes, then tens or hundreds of milliseconds is certainly reasonable (Wood, peer commentary, Libet, 1985, 557)

But can this identity relationship be maintained? If the readiness potential just is the conscious intention, then it must be the case that when the readiness potential is present the conscious intention is also present. Otherwise, it makes no sense to identify one with the other. But this isn't the case: readiness potentials also precede intentions of which we are not conscious (Gomez 1998; Haggard & Libet 2001).

Since the readiness potential can be present without the conscious intention being present, the readiness potential cannot be the same thing as the conscious intention.

Another objection might be that the readiness potentials are not at all causally relevant to our action production. One might argue that all the experiment shows is that something happens in the brain before conscious intentions play their normal role of causing actions. A person holding this view could also claim that no direct evidence has been given for the causal efficacy of these readiness potentials. We do not have evidence for the claim that the readiness potentials cause actions in the way that we have evidence in normal billiard ball examples of causality. Such a person could then claim that, lacking this evidence, we should not be so quick to dismiss our common sense notions about the causal efficacy of our conscious intentions.

Although this objector would be right in claiming that we cannot demonstrate the direct causal chain between the readiness potential and the flicking of the wrist muscles, the arrangement of our brains does lend credence to the claim that the readiness potentials are more likely to cause action than anything in consciousness. In Libet's experiment the electrodes used to detect brain activity were placed at the centerline vertex and the left side of the head (the side opposite the wrist being flexed) (Libet 1998). This means that the readiness potentials are occurring around the supplementary motor area (SMA), primary motor cortex and the pre-motor cortex (Kristeva, Cheyne, and Deecke 1991), all known locations of action selection and production (see, for example, the standard text by Kandel, Schwartz and Jessell

2000). So although we have no direct causal evidence of action production by the readiness potential, given its location and the fact that it occurs first it does seem more likely to be the place where the action is produced.

The arrangement just described should alleviate a third worry, that action production may be overdetermined. One may argue that both readiness potentials and consciousness can cause us to act. More will be said about this in Chapter 3 but suffice it to say that one wanting to hold this view would only have our own phenomenology as a guide, no direct empirical evidence. More to the same point: the brain is divided, for the most part, functionally. Different regions of the brain have different jobs to play and at a certain level the brain isn't set up such that two regions perform exactly the same job. So, if one wants to hold that two separate regions of our brain can control action one needs very compelling empirical evidence that overdetermination can occur in our brains.

I am more than willing to concede that it feels as though we have conscious control of our actions, but the empirical evidence and arguments, to be given in chapter 4, will hopefully demonstrate otherwise. All of the empirical evidence I will present in chapter 3 will point to the will as being solely non-conscious. This idea of our actions being overdetermined seems to stem from a desire to retain some semblance of causal power for consciousness, but lacking any evidence that consciousness does overdetermine intentional action the burden of proof is on those wishing to maintain that consciousness has this kind of power.

The next objection I would like to address here, again from the peer commentary (Naatanen), addresses an issue brought up by other peer commentators (Latto, Ringo). This objection claims that the experiment Libet performed cannot possibly show that, as they put it, "our will" is causally inefficacious, because the experiment does not feature a genuine instance of the exercise of free will. Because the subjects are told by the experimenters ahead of time that they are going to be acting in a certain way, the only choice left to them is when they are going to act in that way. Even this choice is constrained by the experiment. Subjects had to form the intention to act while in the experimental apparatus. "Consequently it appears to be somewhat questionable to describe this motor act as 'spontaneous' or 'fully endogenous' and occurring with 'no preplanning'" (1985, 549). But this objection makes a mistake in assuming that you can exercise your free will only if your choices are completely unconstrained. Even if we grant that the instructions given by the experimenter constrain the subject's choice, the subject is still free, from moment to moment, to choose to act at that moment or not. One could ask if most of the choices we make in day-to-day life are free in the way that Naatanen seems to require. For example, if my boss tells me that I can take lunch anytime between 12:00 p.m. and 2:00 p.m. but that I must take a lunch break, is my choice of taking a lunch break from 12:30 p.m. to 1:30 p.m. not an example of freely choosing? If it isn't, then according to Naatanen, a significant number of our choices are not freely chosen. But this is deeply counterintuitive. If my choice of lunch time was unconstrained then so were the choices made by subjects

involved in the experiment. Naatanen is setting too high a standard for genuine free will.

Having dealt with these objections I hope I have made Libet's position clearer and more plausible. As mentioned, there is much more to say on this in the final chapter. But, hopefully it is evident that there is a real puzzle here, that we have a case, either of our phenomenology misleading us into thinking we have control when we don't, or of some empirical evidence that needs to be explained away. I, of course, will choose the former part of the disjunction that the empirical evidence is right and we are wrong. But before doing so I would like to examine some theories of the will and see what they have to say about the will. After all, what's an argument without someone to argue against? But I will also present one view that complements mine, that provides further empirical evidence to back up what I'm claiming. As much as I hope this chapter has introduced and clarified what I'm arguing about, I hope the next chapter will present the larger picture: what other people who talk about consciousness and the will have to say about the issue.

Chapter 2

In what follows I am going to introduce a number of theories that, in one way or another, attempt to explain the relationship between the will and consciousness. Three of these theories (those of Descartes, Searle and Strawson) are attempts to show that the will and consciousness are in some sense identical. The last theory (Wegner) is an attempt to prove the opposite: that the will and consciousness are separate entities and one isn't necessarily required for the other. The reader will have to wait for Chapter 5 to discover how I am going to deal with these theories but what follows should lay out the groundwork for some important theories of how the will and consciousness are related.

Descartes

In his various writings Rene Descartes says many things about the will. In *The Passions of the Soul* Descartes clearly explain what he takes to be the relationship between the will and consciousness. It is here that Descartes claims that our feelings of acting willfully just are what the will is.

For it is certain that we could not will anything unless we perceived that by the same means that we willed it. And though with respect to our soul it is an action to will something, it can be said that it is also a passion within it to perceive that it wills. *Nevertheless, because this*

perception and this volition are really only a single thing, the denomination is always made by the loftier one, and so isn't usually named a passion, but an action only. (Descartes 1989, 29; italics my own)

Here Descartes seems to be making the claim that our feelings of volition and our volition are the same thing. So when I attempt to open a particularly sticky door the feeling I have of trying to exert pressure on the door is just what it is to act willfully to open the door. The feeling of willful trying is just what the will is. If this is the case then it seems as if the will has to be part of what we would call 'consciousness' and Descartes would call the soul, since feelings exist in the soul/consciousness and nowhere else. So it seems that for Descartes there can be no separation of the will and consciousness: the will must be part of consciousness since to will and to perceive that we have willed are the same thing and perception must surely occur in the conscious part of us.

Searle

John Searle has many reasons for arguing that the will and consciousness are, in some way, related. Searle holds that intentional states (states having content) have to be possible contents of consciousness. So, all of the beliefs, desires and experiences you have must be able to be part of consciousness. I'll begin by briefly laying out Searle's

most recent theory of agency and then go on to show why consciousness is required, by Searle, for willing/volition.

For Searle there is a question that needs to be answered about our actions. Searle thinks there is a gap between our beliefs and desires and the actions we perform. The beliefs and desires are not, according to Searle, causally sufficient for producing action. What he means is that given the same total set of beliefs and desires we could have chosen to perform some other action than the one we actually chose. This is what Searle refers to as "The Gap" and the question he aims to answer is "What fills the gap between our beliefs and desires and our actions?" Searle claims that the thing that fills the gap is the self and then goes about assigning certain properties to the self:

The features of the self can be stated as follows:

There is an x such that

1. x is conscious.
2. x persists through time.
3. x operates with reasons, under the constraints of rationality.
4. x, operating with reasons, is capable of deciding, initiating and carrying out actions, under the presupposition of freedom.
5. x is responsible for at least some of its behavior.

(Searle 2002, 95)

Since all of these characteristics are necessary for the self, consciousness must be present for action to be produced. So on Searle's view willful actions require consciousness. But, for Searle, it's not as if the will resides in consciousness, it's simply that, as Searle puts it, "every unconscious intentional state is at least potentially conscious" (Searle 1992). That is, every perception, thought, wish, belief, desire, intention – every mental entity which is about something – is one which is potentially a part of consciousness. Searle thinks this is true because our current non-conscious mental states lack the fine grained "aspectual space" (something like Fregean sense [reference]) to distinguish between co-referential mental states. So when we think of water we think of different things than when we think of H_2O and it is this difference that makes our water thoughts about water and our H_2O thoughts about H_2O . The following should help clarify what this difference is.

Both water and H_2O will cause me to have the same waterish conscious experience, because water is H_2O and everything having the former property will have the latter as well. What makes my water-thoughts differ from my H_2O -thoughts - what gives them different aspectual shapes - is their respective structure, not my visual experience of a transparent or murky (as the case may be) looking liquid, nor my conscious "gustatory" experience of the "tasteless" liquid dripping on my taste buds and then flowing down my throat. What makes my water-thoughts differ from my H_2O -thoughts is that

the latter, unlike the former, involves my thinking about hydrogen, the number 2 and oxygen. (Jacobs, 1995, 195-200)

Aspectual shape must matter to the agent. It is . . . from the agent's point of view that he can want water without wanting H₂O (Searle 1983.)

So, aspectual shape is what is required to distinguish my thoughts about water from my thoughts about H₂O. If it were not for the fine grained detail of consciousness my thoughts about water and my thoughts about H₂O would be the same thoughts. So there would be no difference between intending to get some H₂O and intending to get some water. Presumably Searle finds this possibility, that thinking about water and H₂O amount to the same thing, untenable or implausible so has argued for the case that consciousness is what distinguishes these thoughts. Thus it follows, on Searle's view, that every mental state capable of these fine distinctions is capable of being part of consciousness.

But this means that the mental states that make up the will (desires, intentions...etc.) can be in consciousness since we can will to drink some water without willing to drink some H₂O, for example. So, although, for Searle, the will does not always reside in consciousness, there is a built in, necessary relationship between the will and consciousness. If we didn't have consciousness then our desires and intentions wouldn't have contents and therefore wouldn't be directed at anything in the world,

something that is required if they are to be propositional attitudes at all. So in a very real sense, consciousness is required for intentional action because consciousness is the thing that makes it the case that our mental states have content and having no content entails no intentional action.

Strawson

Galen Strawson holds a view on consciousness similar to Searle's. He believes, like Searle, that for our mental states to be intentional they have to bear some relationship to consciousness. So, like Searle, for Strawson consciousness is necessary for the states that make up the will to be about things in the world. The other motivation to look at Galen Strawson is that some philosophers, for example Frankfurt (1998), would think that "the will" is the name of some sort of desire. Since Strawson has a lot to say about desire, his theory, some might claim, implies something about the will. So, it is important to examine what Strawson says about the nature of desires and their role in our conscious lives.. First I will explicate Strawson's views on intentionality and consciousness and their similarity to Searle's views and then I will move on to talk about Strawson's views on desire.

Strawson, like Searle, holds that consciousness is required for intentionality. He thinks this primarily because he thinks that experience is what makes our concepts about certain things. He thinks that primarily because he recognizes the possibility of building machines (robots) that could do everything mental that we could do, that

would be functionally identical to us, without their concepts being about anything because they don't have consciousness. So experience is what makes the difference between us and machines that may be able to do almost anything we can do, but lack intentionality.

A question remains about the place of experience. It seems we can build experienceless machines that can do all the things that we may at first think of as requiring intentionality. They can track objects doggedly through space. They can be very good at distinguishing water (e.g.) from anything that is not water. They can pour forth the decimal expansion of π . And each of these designed machines has an exact and undesigned fluke counterpart that can do anything that it does. For many, it is the undesigned counterparts that drive home the intuition that genuine intentionality is essentially bound up with experience. When an undesigned machine possesses a sample of liquid and prints 'This is (not) water', it seems implausible to say that it is exercising the concept of water. That is the intuition. (Strawson, G. 1994, 203)

So for Strawson it is experience, and therefore consciousness, that makes the difference between intentional and non-intentional mental states. As with Searle, since consciousness is a requirement for intentionality and we think that, at least, some of the mental states that make up the will are intentional, consciousness is a requirement for intentionality.

Strawson's main views about desire come out in his discussion of, and arguments against a position he calls Neobehaviorism. Strawson holds that most people in the philosophy of mind are committed to some form of Neobehaviorism. This theory holds that our mental lives are essentially linked to actual overt, publicly observable behavior. Strawson goes on to claim that he will also argue against the weaker, but more plausible, thesis that our mental states are connected to dispositions to behavior. This thesis is slightly different than the one above but gets lumped in with Neobehaviorism by Strawson. People who hold that our mental states are essentially linked to our dispositions to behavior generally hold the dispositional view because they think that paralyzed, or similarly incapacitated people, still have mental lives even though they cannot behave in any overt way, something that seems to be denied by the stronger thesis.

Obviously the least contentious candidate for Neobehaviorism is going to be desire, since we most commonly associate desire, as opposed to beliefs or representations, with producing behavior. If we could demonstrate that desire, so frequently associated with behavior was, in fact, not necessarily a cause of behavior (or a disposition to behavior), then perhaps our Neobehaviorist claims about the rest of our mental states should be modified as well. This is part of Strawson's aim, to show that Neobehaviorism is false for desires. Since desires are central to the will in a way that beliefs are not I will focus here on Strawson's arguments about desire.

Strawson sets out to prove the Neobehaviorists wrong by imagining a world, the Weather Watchers' world, in which there are beings that have no ability to act overtly in any way. They are firmly planted in the ground and can only observe the world around them, not interact with it.

The Weather Watchers are a race of sentient, intelligent creatures. They are distributed about the surface of their planet, rooted to the ground, profoundly interested in the local weather. They have sensations, thoughts, emotions, beliefs, desires. They pose a conception of an objective spatial world. But they are constitutionally incapable of any sort of behavior, as this is normally understood. They lack the necessary physiology. Their mental lives have no other-observable effects. They are not even disposed to behave in any way.
(Strawson, 1994, 254)

As Strawson says, the Weather Watchers lack any dispositions to behave in any way, but they also desire certain states of affairs. So if the Neobehaviorist wants to maintain some semblance of her position then she must try to find a way of retaining the disposition to act for the Weather Watchers. Strawson argues that it is incoherent to talk about something that cannot act (Weather Watcher) having a disposition to act. This is because conditions that support this kind of disposition would have to posit agency for the Weather Watchers. The conditional that would make it such that the weather watchers were disposed to act would be something like "Were the weather

watchers agents then they would be disposed to act"(Strawson, 1994). But, as Strawson rightly points out, this disposition crosses an important line, a line that would allow anything to be disposed to act.

To say that W (weather watcher) is now disposed to act in certain ways because of the truth of this conditional is like saying that a lump of plastic is now disposed to conduct electricity because its constituent parts could be reorganized to constitute gold. It's like saying that a wet piece of clay is now fragile because it is now such that if it were treated in a certain way (baked in an oven), it would be fragile."

(Strawson, 1994, 275)

One may be wondering what all of this has to do with us. After all, the Weather Watcher's world is merely a postulated world containing implausible aliens and one that doesn't seem very close to the world we reside in.

But there are reasons to think that the Weather Watchers world is closer to ours than, on first blush, it may seem. There are at least two ways in which the Weather Watcher's lack of ability or disposition to act maps onto people in the actual world. There are paralyzed people in our world who obviously lack the ability to act. These people are similar to the Weather Watchers in that we all think that they have desires

(most obviously the desire to be able to act) and yet they lack the ability to act on those desires.

Strawson also urges us to consider the case of a paralyzed person who has been paralyzed for so long that this person has lost any dispositions to act. This person would, in essence, have given up on action in that he would have been paralyzed for so long that he could no longer conceive of himself as an agent. Strawson argues that these kinds of people are prime examples of Weather Watcher like beings existing in our world. Whether or not you find this last argument convincing it should be clear that the example of the Weather Watchers isn't some distant implausible counterexample that bears no relation to humans. We sometimes have the same concerns about our brethren as we do about the Weather Watchers.

The Weather Watchers cannot act and are not disposed to act in any way. But they do have the capacity to desire certain states of affairs to come about. They can desire the sun to shine, they can desire it to rain tomorrow so that their favorite plant will grow, all of this keen interest in the weather is stipulated in the Weather Watchers example. What Strawson wants to claim is that this story is perfectly intelligible, that there is no good reason to suppose that the story of the weather watchers is impossible. And if it isn't impossible then it is possible. If it is possible that there exists beings with desires but no overt behavior or dispositions to behavior then there is no necessary connection between desires and behavior. Hence, the neobehaviorist claim that there

is a necessary link between desire and behavior is false. This is Strawson's main argument against neobehaviorism.

So what happens as a result of the desires of Weather Watchers if they don't act or are not disposed to act in any way? The first part of the answer is that they philosophize; they introspect on their mental states and make assessments of what is going on in their own minds. They notice and analyze things like the fact that their desires tend to occur at the same time as certain other mental states. The second part of the answer is that they receive pleasure or displeasure from having their desires satisfied or frustrated. Having desires satisfied makes the weather watchers happy and having them frustrated makes them unhappy. Strawson claims that this property of desire, that desires dispose us to pleasure or displeasure, is the essential feature of desiring, not some connection with behavior.

As remarked - and leaving aside all facts about the phenomenology of conscious episodes of desire, important as they are - the primary linkage of the notion of a desire to a notion other than itself isn't to the notion of action or behavior but rather the notion of being pleased or happy or contented should something come about (or at least to the notion of ceasing to be unhappy or discontented should it come about) and to the distinct but correlative notion of being unhappy or discontented should it not come about. (Strawson, 1994, 280)

So instead of a connection between desire and action we are instead to hold that desires have an intimate connection with pleasure and displeasure. One might think this is something like psychological hedonism. But it differs from psychological hedonism in that Strawson is claiming that the end result of desiring is pleasure or displeasure not that the only thing we desire is pleasure. So pleasure is not the object of our desire, according to Strawson, it is the result of having our desires satisfied.

Strawson's account may seem counterintuitive. This account of desire doesn't seem to capture all the complex emotional combinations that we seem to experience as a result of desire satisfaction or frustration. If we take the case of a person who feels elation, say at the discovery that her book is getting published and at the same time deep emotional turmoil, as the result of finding out that a parent has passed away, how do we account for these complex emotions under Strawson's theory? The answer is that there is no limit on the number of standing desires one may have at any given time and that any of these desires may cause their respective affective response depending on whether they are frustrated or satisfied.

So for Strawson there must be some sort of weight each of these satisfied and frustrated desires contributes to our current affective state. Our desire to be published might cause an increase in pleasure that is outweighed by the frustration of our desire that our parents stay alive, leading to an overall negative feeling. Strawson isn't required to give a full account of how our desire frustration and satisfaction is to be weighed and combined. It is enough that this view seems intuitive, at least to the

extent that we think that desire satisfaction and frustration do contribute to our feelings of pleasure and displeasure.

Also, Strawson lays no claims on what type of pleasure or displeasure must occur as a result of our desires being satisfied or frustrated. Nor does he place any limit on the number of desires that can simultaneously be frustrated or satisfied by how we perceive the world. So complex emotional reactions are possible under Strawson's theory just because there are associated complexes of weaker and stronger desires that can simultaneously be satisfied or frustrated at any given time. Were it the case that the woman above had a weaker desire to get her book published, she may have been pleased instead of elated. Had she no regard for her parent's well being she may have felt nothing upon hearing that her parent had passed away but still be elated by her publishing deal. She could also have a third desire for a glass of water that could add further complexity to her current emotional state. All of this is explainable, within Strawson's theory, because of the underlying desires.

It may seem as though Strawson has left us with an incredibly vague account of how desires dispose us to pleasure and displeasure. We have been given no account of how to quantify desire strength or our responses to their satisfaction. We haven't even been given a way to correlate a desire with certain responses. Strawson has left things this open just because there are complex emotional reactions to a complex world that we need to be able to account for. In the case of the woman who feels both elation and turmoil at the same time, how could we account for these odd but real emotional

responses if it were not the case that we had a somewhat loose theory of desire satisfaction that would be amenable to these complex responses? Sometimes people really do laugh and cry at the same time and an overly rigid account of how desires lead to pleasure and displeasure will have a difficult time accounting for this.

Since, according to Strawson, all desires have this disposition to cause pleasure and displeasure in us and since pleasure and displeasure, happiness and sadness, joy and pain are all experiences and therefore conscious states (according to Strawson) then it seems as if desires have to be in consciousness. For this reason, and the one previously mentioned, Strawson and Searle hold similar views. Searle holds that consciousness is required to give desire and the like intentionality. Strawson seems to be committed to the claim that consciousness is required because that is where pleasure and displeasure take place, but also that experience is required for intentionality. For both there is something that consciousness gets us that they take to be an essential feature of mental states (at least the ones with intentionality) so consciousness must be present for those mental states to have the right characteristics.

Wegner

In his book, *The Illusion of Conscious Will* (2002), Daniel Wegner does critical research into psychological aspects of the will and its relationship to consciousness.

His end conclusion is that conscious will is illusory, that we only feel as though we are in conscious control of our actions when, in fact, we are not. Wegner goes about this, for the most part, by pointing out situations in which we lack conscious will and then explains the relevant underlying process in a way that motivates his conclusion. Of course, in any lengthy research project some of the evidence is better and some worse. In what follows I just want to highlight some of the key studies that Wegner gives as evidence that our feelings of conscious will are illusory⁵.

Alien hand syndrome is possibly the most fascinating piece of the empirical evidence talked about by Wegner:

Alien hand patients typically experience one hand as acting autonomously. They do not experience willing its actions and may find it moving at cross purposes with their conscious intentions.... Banks and colleagues (1989, 456) report an alien hand patient whose 'left hand would tenaciously grope for and grasp any nearby object, pick and pull at her clothes, and even grasp her throat during sleep...She slept with the arm tied to prevent nocturnal misbehavior. She never denied that her left arm and hand belonged to her, although she did refer to her limb as though it were an autonomous agent.' (Wegner, 2002, 4)

⁵ Of course, part of Wegner's evidence is the Libet experiment previously mentioned in Chapter 1, so I will exclude that in the discussion to follow.

What are we to think about the alien hand case? As Wegner points out, the hand seems to act purposefully; it has the ability to carry out actions that seem to have intentions guiding them. Yet the patient cannot stop the hand from acting and does not consciously initiate the action. So does this demonstrate that consciousness isn't necessary for willful action? I think it does, in that there doesn't seem to be any other way to account for the alien hand. There must be some non-conscious will guiding the hand. After all, the hand does not flail about or only act in one way, as one might expect from some random or autonomic behavior. The hand acts as though it was being directed by the will of an owner, and yet the actual owner cannot consciously control the hand. Therefore there must, at least in alien hand patients, be some separation between our will and consciousness.

Wegner also addresses a point that will be brought up again in more detail in Chapters 3 and 4, that there is a separation of the parts of the brain that give rise to consciousness and the parts of the brain that control action. Wegner begins by discussing the relevant parts of the brain associated with action production, namely the frontal lobes. In MRI and EEG studies the frontal lobes have been implicated in action production. Along with this, people who suffer specific types of frontal lobe damage lose the ability to plan for and perform actions "This widely observed phenomenon suggests that many of the causal sequences underlying human action may be localized in brain structures just under the forehead." (Wegner, 2002, 30).

Wegner goes on to introduce the studies of Wilder Penfield and his famous experiments on conscious patients and their experiences during surgeries in which he stimulated portions of their brains while they were conscious. Stimulating certain parts of the brain could cause the patients to act in certain ways, ways that were elaborate and seemed intentional, but because the patients were conscious they were able to report that they had not caused the actions to happen. Instead the patients claimed that Penfield made them do it or that they didn't want to act but that they did anyway. So again, willful actions can be produced without conscious intent.

Contrary to Penfield's work is the work of neuroscientist Jose Delgado. Delgado performed similar brain stimulation studies to Penfield but found that his patients did feel that they had initiated the action. But Delgado believed that this feeling of volition reported by the patients was merely a way for them to justify the fact that they had acted or may have been a hallucination of the patient. Wegner interprets the Penfield and Delgado studies in the following way:

No matter where such an area might be, though, the comparison of Delgado's patient with the one examined by Penfield suggests that the brain structure that provides the experience of will is *separate* from the brain source of action. It appears possible to produce voluntary action through brain stimulation with or without the experience of conscious will. This, in turn, suggests the interesting possibility that conscious

will is an add-on., an experience that has its own origins and consequences. (Wegner, 2002, 31)

The rest of Wegner's book is a walk through some interesting but non-definitive evidence for the separation of consciousness and the will. I am not sure of the value of the research, but there is reason to talk about it anyway because there is a lot of evidence and in the absence of some hard empirical research (see Chapter 3 for that) a lot of interesting anecdotal evidence can still indicate something, if not prove it demonstrably.

Wegner talks about a lot of related phenomena that I want, for the sake of space, to lump together. These I will call the 'loss of self' phenomena. These include things like hypnosis, spiritual trances, sleepwalking, "possession" and the like. Even though there are obvious differences between these cases there is at least one common feature as well, and it is this common feature that Wegner wants to draw our attention to. In all of these cases we lack the feeling of conscious will and yet we are still able to act purposefully. We are not in conscious control of ourselves and yet purposeful, seemingly willed action is being produced. In the case of hypnotism it seems likely that the willed action is the result of instructions by the hypnotist. For spiritual trances and possessions perhaps, according to Wegner, we hypnotize ourselves and induce some kind of global mental state that lacks consciousness but controls action. Whatever the explanation for these phenomena they all point to a separation between

the conscious feeling of will and being able to act intentionally, because willed action is present and consciousness isn't.

Chapter 3

In this chapter I plan to examine the neurophysiology of action selection and production. What I hope to show at the end of this is that there is a specific place in the brain where actions are selected and initiated and that this part of the brain isn't one associated with consciousness. I take it to be the case that the will is what gets us to act intentionally. If there is one specific part of the brain that performs this function then it seems to me that this is where the will must be. So, if we can point to a region of the brain as the initiator of willed actions and that part of the brain isn't part of consciousness, then it follows that the will isn't in consciousness.

One might wonder what I mean when I claim that certain areas are the initiator of intentional action and some are not. To be clear, what I mean is that these areas of the brain are the primary and original causes of our willed actions. Were these specific areas of the brain not present, there would not be willed action. Of course we could still act; we would still have a spinal cord and neurons to fire it. But, lacking these brain regions, none of our actions would be willed, they would only be reflexive they would not reflect any choice on our part.

A Walking Tour of Action Selection/Initiation

Before one begins a walking tour of anything it is important to have a map. One wants to know what stops there are along the way and what each of these stops has at it. Usually the map will give information about various activities that go on at those locations and will provide details as to how to get to other locations on the map. So let me start by detailing the areas of the brain we'll be visiting and certain interesting features of them. Once that is done we'll go through all of the areas in sequence, in the same sequence that information takes when intentional action initiation actually takes place.

Motor Areas of Cortex

The motor areas of the cortex are the Supplementary Motor Area (SMA), Pre-motor Cortex (PC) and the Primary Motor Cortex (PMC) (KSJ, 757, 1991). The motor areas of the cortex take in information from the senses and beliefs. This information is then used to generate a number of various action plans (Mink 1996). These action plans reside in the motor areas of the cortex until certain information is sent from the basal ganglia (more on that later). The motor areas of the cortex are sometimes referred to as the homunculi, the little guys that live inside of us that have parts corresponding to our parts and decide how parts of our body are going to act. Except for the decision part I take this picture of the motor cortex to be fairly accurate: there are areas of the motor cortex that are related to movements in specific parts of the body.

The motor areas of the cortex are connected to the spine, the basal ganglia via the thalamus and the somatosensory cortex (KSJ, 757, 1991) as well as a few other areas that are uninteresting for philosophical purposes. The projections from the motor areas of the cortex are primarily glutaminergic (the main neurotransmitter used by the brain to excite the receiving neuron into action). As noted the motor areas of the cortex receive information from the senses and our beliefs via various channels. All of this information comes together in the motor areas of the cortex and allows the cortex to make action plans that take into account our surroundings, including how our body is behaving right now and what we desire⁶.

Another interesting fact about the primary motor cortex: It always fires before action is produced (KSJ, 858, 1991). So every time we act the primary motor cortex has some role to play in that action. This might make the primary motor cortex seem like an excellent candidate for the seat of the will, but it is important to keep in mind that not all of our actions involve the will. If the primary motor cortex is the location of the will then there is no distinction to be made between willed action and non-willed action. Since there is such a distinction, the primary motor area cannot be the seat of the will.

Basal Ganglia

This is where things get tricky. There are many subdivisions of the basal ganglia and not every subdivision plays a role in action. Only some of these sections are directly relevant to action selection but others have influence over this decision. Before subdividing the basal ganglia let's take a look at its general structure.

The basal ganglia may be viewed as two primary input structures, two primary output structures and two intrinsic nuclei. ...The input structures of the basal ganglia are the striatum (caudate and putamen) and the subthalamic nucleus. ... The output structures are the globus pallidus internal segment (GPi) and substantia nigra pars reticulata (SNpr). ... The intrinsic nuclei are the globus pallidus pars externa (GPe) and substantia nigra pars compacta (SNpc) (Mink, 1996, 383)

As noted some of these areas offer input to regions of the basal ganglia as a way of influencing which action is selected (Schroeder 2004). So not all of the basal ganglia is directly related to action selection in the sense that some areas influence action selection and others directly project signals that select certain actions. So the basal ganglia both makes decisions about action and issues commands about how to act. But, much like how a gallery that has a café in it is still a gallery, the basal ganglia is seen primarily as a motor control area (Mink, 1996, 382). Perhaps then we should start by subdividing the basal ganglia into motor and non-motor areas. For the sake of clarity I will focus primarily on the GPi, the GPe, the striatum, the SNpr and the subthalamic nuclei (STN).

⁶ According to Schroeder (2004) our desires are located in brain regions that are also part of the will, the (SNpc) being the most obvious example. So when we are deciding how to act the region of our

These areas are also interconnected. The GPe sends inhibitory output to the STN, GPi and SNpr. The STN sends excitatory output to the GPi, GPe, which also receive inhibitory input from the striatum. The GPi is the primary output back to the motor cortex, via the thalamus. (Mink, 1996, 383)

Now that we have a better understanding of the layout of the motor control sections of the brain, it is time to begin the tour. As with all tours, not everything is on the map (if it was, why take the tour?). So I'll be adding information along the way, such as the type of connections between areas and some more specific information about certain regions. What I hope to capture here is the path that information takes in your brain as you decide to act.

The Tour

We begin in the cortex, where information from the sense modalities is collected and, probably, beliefs are processed. This information is then used by the motor areas of the cortex to make what some people (Jeannerod, 1997) call "action scripts" (or Mink calls "motor program generators"). These are like real theatrical script in that each contains information about how to act. But instead of stage directions we have action

brain that decides is directly influenced by our desires.

directions. And much like a theatrical warehouse there are many scripts residing in motor areas of the cortex. So, like a million monkeys at a million typewriters these areas of the cortex churn out script after script in the desperate hope that one will be acted on. One might wonder why it is that we generate so many action scripts when only one is ever needed for each action. This is because there are many potential actions that could be performed at every turn.

These action scripts, or information about them, are then fed to the striatum. Each of the three main areas of the motor cortex all project to a different area of the striatum. So the SMA, PMC, and PC project to three different places. For this reason the striatum is seen as a focusing area for information from the cortex. "The compartmental organization and pattern of termination of cortical inputs provide an anatomical basis for focusing information in the striatum." (Mink, 1996, 385) Areas of the striatum are selectively inhibited and excited by the SNpc using two different types of dopamine receptors (D1 and D2). Since the SNpc is primarily a reward projection center (Schroeder 2004) it seems plausible that at this level what is happening is that information coming down (yes really down) from the cortex to the striatum is then "pruned" at the striatum. By which I mean that, at this point action plans that do not meet with or satisfy any standing desires may be cut. If this is the case what seems to be happening is that certain scripts are automatically selected for dismissal. Think of it as the first round of editing for a large pile of scripts, some get sent on for further editing, some get tossed out as un- producible. Those selected as passable are allowed to go on to the other regions of the basal ganglia.

The next stop on our tour (apparently of the theatre district) involves the areas that seem critical for action selection. From the striatum the information that got through goes to both the GPe and the GPi via, primarily, the GABA neurotransmitter (Mink, 1996) (The GABA neurotransmitter limits the firing of other neurons, performing the opposite role that glutamate does). The information from the striatum is also passed on to the STN. The GPe is excited by the STN and itself inhibits the STN. (Mink, 1996) The GPe also inhibits the GPi and the STN via the neurotransmitter GABA. So at this level we have a number of areas of the basal ganglia inhibiting and exciting one another. But it is difficult to say exactly what is going on where. Does the information from the striatum stay in the GPi and then get augmented by the GPe and the STN? Does the GPi send information to either the STN or the GPe? This is unclear. Wherever the information goes within the basal ganglia it clearly has only one way out. Only the GPi and SNpr project back to the motor areas of the cortex via the thalamus. This might suggest that the information from the striatum resides in the GPi and is acted on by the STN and GPe until it is released.

Although it is difficult to characterize the function of and interaction between the GPi, GPe and STN, what can be said is that the effects of the GPi's projections back to the motor areas of the cortex is that it inhibits most action scripts and selectively disinhibits the one that will be acted on (Mink, 1996). The motor areas of the cortex, at this point, are like Broadway. The script has been selected and all others are in a filing cabinet in some agent's office. The GPi has selected one action in disinhibiting

it, while all others are being inhibited. At this point the play starts as information is finally sent out to the spine to cause our bodies to act. This is the end of the tour, please exit the bus in a calm and orderly fashion.

It should be noted that although I have run through the action selection process in a sequential, quantified (single action) manner, this whole mechanism acts in a fluid manner, as seems obvious from the fluidity of our everyday actions. Also some of these mechanisms do run in parallel. The STN is inhibiting something in the GPi at the same time as the SNpc is helping to focus information in the striatum and a number of the other processes mentioned run in parallel.

One might be wondering how one action gets selected over another: what's the exact mechanism that allows one action to be chosen and another not. A few theories have it that the GPi is like a theater critic; it pans some scripts and praises others. One theory (Humphries and Gurney, 2002) claims that the GPi selectively ramps up the salience of certain action scripts and makes others less salient. So once the information leaves the GPi one script has so much more salience than any other that once it travels back to the cortex it is the script that gets acted upon. It does this, according to the theory, in conjunction with the STN and the thalamus. This theory, if it is correct (the jury is still out), also claims that this loop between the cortex and the basal ganglia is the only area of the brain that has the right functional and computational layout for action selection. According to Mink (2002) the mechanism by which selection happens is that the GPi inhibits all action scripts but it selectively

releases this inhibition from specific action programs, so the mechanism for selection is to inhibit everything that shouldn't be acted on more than the one thing that should be. This will all be made to seem more plausible as we move along. However this is just the mechanism by which action gets selected, it doesn't give us a clearer picture of why or how one selection is taken over another.

Interestingly enough, although there is consensus in the scientific community that action selection takes place in the basal ganglia there is still no good theory as to what the exact mechanism for that selection is. It is as if a black box resides in the GPi (or the loops between the GPi, GPe and STN). We can record activity in that area of the brain but God only knows what that information represents. If we see a spike in activity in that area of the brain we know that something is going on but it could be the sending of information to the STN just as easily as it could be some operation to select action. What we do know is that, taken as a whole, the basal ganglia does look like the area of our brains where action is selected. This isn't just conjecture, nor is it based on some computational model of how we think action selection works. In addition to the physiological evidence, further evidence for the basal ganglia being the location of action selection is that injury or damage to the basal ganglia results in a lack of ability to select and initiate action, especially voluntary action. In what follows I will look at some of the evidence suggesting that the basal ganglia must be where action selection takes place.

The following is some of the primary brain injury evidence that has convinced people to see the basal ganglia as the seat of the will: "In general, hypokinetic (e.g. Parkinson's disease) or hyperkinetic (e.g. Huntington's disease) movement disorders are associated with increased or decreased basal ganglia output, respectively" (Wichmann and DeLong, 1996). The explanation goes like this. In Parkinson's disease the GPi is disinhibited by the STN (or the death of too many SNpc cells, which project dopamine). This means that the GPi acts more than it normally would, but the action it primarily takes is to inhibit action scripts in the cortex. So now that it is disinhibited (acting more) it inhibits more actions in the cortex (this effect is being mediated by the GPi's effects on the thalamus). This makes it harder for any one action to be released from the cortex to the spine, hence the effect of severe Parkinsonism: people have a very difficult time getting their bodies to act how they want. With Huntington's disease the opposite is true, the GPi is inhibited (acts less), so it inhibits fewer actions scripts in the cortex (more scripts are sent to the spine) than it normally would. This causes conflicting actions to be produced in the body (imaging flexing your bicep and triceps at the same time) which results in rigidity as the muscles both try to move the same limb in opposite directions.

The results of anatomical, physiological and lesion studies can be integrated into an overall hypothesis of what the basal ganglia contribute to movement. This hypothesis states that the tonically active inhibitory output of the basal ganglia acts as a "brake" on motor pattern generators (MPGs) [What I have been referring to as "action scripts"] in the cerebral cortex (via thalamus) and brainstem. When a movement is

initiated by a particular MPG, basal ganglia output neurons projecting to competing MPGs increase their firing rate, thereby increasing inhibition and applying a “brake” on those generators. Other basal ganglia output neurons projecting to the generators involved in the desired movement decrease their discharge, thereby removing tonic inhibition and releasing the “brake” from those desired generators. Thus selected movements are enabled and competing postures and movements are prevented from interfering with the one selected. (Mink, 1996, 414)

Further evidence of the role of the basal ganglia in action production comes from the very setup of the loop between the cortex and the basal ganglia. If the basal ganglia did not play an important role in action selection then it would be a mystery why motor regions of the cortex project to it so strongly. Also, why does there seem to be less information coming out of the GPi than there is going into the basal ganglia from the cortex if the basal ganglia doesn’t do anything with this information? These things are a mystery if the basal ganglia doesn’t have a central role in action selection.

So it seems as if we have good theoretical and empirical evidence that the basal ganglia is the location of action selection and initiation. But things get better. Since Mink’s article in 1996, more research has been done on the function and characteristics of the basal ganglia and its role in voluntary movement. The following should be taken as a summary of what many neuroscientists are now willing to claim about the basal ganglia:

What do all these findings mean for the issue of controlling voluntary action? Put simply, it seems that for internally-guided voluntary actions, which are not sufficiently automatized, the pre-frontal and orbito-frontal cortex (in cooperation with other areas) plans, and prepares a certain action. This already happens under strong influence of the unconscious limbic-emotional centers (above all amygdala and meso-limbic/meso-cortical systems) via the thalamic loop. Interestingly, our conscious brain is unaware of the subcortical origin of many, if not most of our wishes; they seem to come from 'nowhere'. Before being eventually realized, these wishes and plans must pass through the dorsal loop for a second subcortical 'censorship' before the cortex can influence the pre-motor and motor centers of the cerebral cortex, so that these determine the action in detail and induce them. This censorship of the basal ganglia is, in turn, controlled via the ventral loop by the experiential declarative and emotional memory contained in the limbic system. (Roth, 2003)

So information about how things are going (via the amygdala and the SNpc) reaches the basal ganglia and this information is processed as a way of computing the best possible action in the current situation. The type of information the basal ganglia appears to compute over in deciding the best action is something like the following:

[cut extra line]

It appears that the censorship of the basal ganglia is essentially concerned with two questions:

1. Whether *in light of previous experience* the planned action should really be carried out.
2. Whether the intended action is *appropriate* to the situation.

Only when the two questions are answered affirmatively do the basal ganglia activate the pre-motor and motor cortex via the thalamus (more precisely, they reduce or abolish the inhibition of the thalamic nuclei). This happens via the 'dopamine signal', which in turn is under the control of the unconscious emotional-experiential memory. (Roth, 2003)

So it would appear that certain areas of the brain have influence over the basal ganglia but that the primary location of decision making about which actions to perform and which not to is only to be found in the basal ganglia. What Roth points out and what is of critical importance for my purposes is that we are never conscious of the decision making in the basal ganglia. Certainly, consciousness (or experiences in consciousness) might influence which action gets chosen. It wouldn't make sense to intend to get in the car if there wasn't a car there. But, at best, consciousness is just one of the many brain regions that input to the basal ganglia and on whose information the basal ganglia might or might not base a decision. So it's not as if consciousness has no role to play in how we act, just that its role is to tell the basal ganglia what is going on so that the basal ganglia can then make decisions about which action is best.

But all of this is moot if the basal ganglia isn't the location of the will. Yet it seems as if this has to be the case. We think of the will as the thing that gets us to act intentionally, that the will is what makes the difference between autonomic action and other actions. If we think of the will as also being located in space that space had better be where our intentional actions come from. Since our actions, at least our intentional, non-automatic actions come from the basal ganglia then that has to be where the will is. But by transitivity if the basal ganglia isn't a part of consciousness and the basal ganglia is the location of the will, then the will isn't part of consciousness.

*Chapter 4*⁷

In the last chapter I introduced a lot of neurophysiological evidence that was supposed to motivate the claim that the will isn't a part of consciousness. In this chapter I want to move away from science and into philosophy (at last!). In this chapter I will provide more philosophical arguments that the will isn't a part of consciousness. But it is important to keep the empirical evidence of the last chapter in mind as the evidence and the philosophy, hopefully, work together to provide a coherent and mutually supportive picture of the relationship between the will and consciousness.

Imagine a woman is sitting on a stool in an ice cream parlour, chatting, and her consciousness teems with diverse elements. It includes her experience of the firmness of the stool and the noise of the crowd, her awareness that her companion is trying to make a joke, her conscious understanding of what was just said to her, her conscious desire to eat some chocolate ice cream and her consciously willing herself not to eat the ice cream.

These various elements of her consciousness can be divided into two distinct categories. As a contemporary philosopher of mind might put it, her consciousness

⁷ The following is a reworking of a paper written with Tim Schroeder and Cynthia N. Read for *American Philosophical Quarterly* (Jan, 2004)

includes both her perceptions (of the stool, the noise) and certain of her propositional attitudes (conscious beliefs about others' intentions and meaning, a conscious desire, an instance of conscious willing). Of course, she has many other unconscious propositional attitudes, and many currently unexercised capacities for experience: her current conscious life does not exhaust her mental life by any means. But her current conscious life is nonetheless representative, one might say. It includes a range of experience-types she is capable of having, and a range of potentially conscious propositional attitudes she possesses.

The description given of this woman is slightly controversial, in that a few philosophers hold that every element of an individual's consciousness must be nonconceptual (e.g., Dretske, 1995, Tye, 1995). If these philosophers are right, then there can be no conscious beliefs or conscious desires – at least in the usual sense, in the sense that there can be no beliefs or desires which are literally elements of one's conscious life – because beliefs and desires, being concept-involving, are necessarily excluded from consciousness. But the view of these philosophers is very much a minority view, although if they are right then so am I. The majority view is that many propositional attitudes can literally be elements of one's conscious life. I am prepared to dispute the majority in general, just in the specific case of the will.

Particularly importantly for present purposes, the majority view, as seen in chapter 2, is that the will can literally be part of one's conscious life, just as perceptions of noise can be. As we have seen some hold that this is a necessary fact. In *The Passions of the*

Soul Descartes gives a typical Enlightenment expression of this view, writing that "...this perception [of willing] is really one and the same thing as the volition..." (Descartes 1988, 226). Also, we have seen that Searle holds that "...we understand the notion of an unconscious mental state only as a possible content of consciousness" (Searle 1992, 155)⁸.

The purpose of this chapter is to give a purely philosophical argument that the mental states cause us to act intentionally, of necessity are *never* elements of one's conscious life: most prominently, instances of willing.

The argument begins by noting that different mental states have different directions of fit, as Searle (1983; 1992), Anscombe (1957), and others have pointed out. In the case of beliefs and perceptual experience, the direction of fit runs from our minds to the world: a belief that P is beholden to the world, and should be changed if the evidence is that not-P. If you are misperceiving or have a false belief your mental state is at fault, not the world; the world is the measure of the correctness of our beliefs. In the case of willing just the opposite is true: the direction of fit is from the world to our minds. Willing that P isn't beholden to the way the world is, but rather puts a demand upon the world. When we will something to happen, we want the world to be a certain way: if our will is effective then the world must have worked in the right way to accord with what we willed, but if our will is ineffective, this isn't a sign that there

⁸ Galen Strawson (1994) has also been quite explicit about the necessary connection he sees between desire and consciousness. For an objection to Searle that applies equally to Strawson, see Van Gulick (1995).

is anything wrong with the will. Sometimes our bodies don't work in accord with how we will. We could be paralyzed (or have Parkinson's or Huntington's or any number of other disorders that interfere with action), but in these cases it isn't some defect of the will, it is a defect of the body in carrying out what the will is instructing it to do.

This way of characterising direction of fit is common but not universally accepted, and other ways of explaining direction of fit also exist. One can, for instance, point out that the notions of accuracy and veridicality apply to perception and belief, but not to the will, to which notions of effectiveness are more appropriate. (This is, perhaps, the most basic direction-of-fit fact, accepted by almost everyone.) One can point out that it is possible to misrepresent but not mis-will⁹. One can point to the fact that a characteristic causal consequence of a mismatch between belief and the world is the adjustment of the belief, while a characteristic causal consequence of a mismatch between what we will and the world is the adjustment of the world. One can point to the distinct biological functions had by truth-directed mental states and action-directed mental states. And so on and so forth. It isn't important, for our purposes, exactly what the difference between the belief-like mental states and the will-like mental states comes to. All that is important is that the difference exists, and can be recognised (if imperfectly) by considerations at least in the neighbourhood of

⁹ In some sense one can mis-will in that one can will oneself to do something that isn't practically rational (punching yourself in the face) or something that is immoral (punching your advisor in the face... probably not practically rational either). But these are better characterized as instances of inappropriate willing. A genuine case of mis-willing would involve willing yourself to act in a way that your will didn't want you to act. I take this to be a contradiction and, therefore, impossible (at least in this world).

those just mentioned. I call this difference one of "direction of fit," as this label is a familiar and convenient one, but I take no particular stance on what the difference consists in.

Since the will has a direction of fit different from perceptual experience, it can be shown that there is no conscious will – at least, no conscious will in any interesting sense, though there is a sensible use of the phrase, to which this chapter will come in a moment. To begin, a survey of the elements of consciousness will be made, considering their directions of fit one by one. A more general argument against any so-called "conscious will" having the direction of fit characteristic of the will will then be made.

To begin: imagine willing yourself to get a piece of chocolate in front of you, willing it in a manner you would describe as having a conscious will. What elements are found in your consciousness under such circumstances? There is your visual perception of the chocolate and perhaps olfactory perception as well. These elements of your consciousness are obviously not your will, as can be shown (if it needs to be shown) by observing that these perceptions obviously have a mind-to-world direction of fit. If you seem to perceive chocolate in front of you where there is none, it is your perception, not the world, which needs adjustment.

In addition to seeing the chocolate and smelling it, perhaps you feel a rumbling in your stomach or anticipatory muscle tension as you begin to reach out for the

chocolate, the latter being commonly called the felt ‘urge’ to take and consume the chocolate. These feelings are also forms of perception – you perceive the condition of your stomach through the nerves in the stomach wall, you perceive your muscle tension through the nerves in your limbs¹⁰ – and so they too have the wrong direction of fit to be the will. The sudden rush of saliva to your mouth, the palpable palpitations of your heart: these too are perceptions you have, and so all share the mind-to-world direction of fit. Just as in the case of seeing and smelling, so here too, if there is a gap between perception and reality, the fault is with perception. An apparent rumble in the stomach might be hallucinated, as might apparent heart palpitations or wetness in the mouth, and hallucination is a problem at the mental end of things, not the world’s.

So an inspection of the contents of consciousness upon an occasion of “consciously” willing yourself to get chocolate reveals nothing with the right direction of fit to be the will. There is one last option those trying to keep the will in consciousness might argue for but we will come to that in the objections section.

One might think that our itemization of the elements of one’s conscious life during an instance of willing oneself to get chocolate is incomplete. One might argue that inner utterances such as “Get that chocolate!” are both in consciousness and are instances of willing. Admittedly, this isn’t a simple matter of direction of fit, for an utterance of “Get that chocolate!” plausibly has the will-like direction of fit, even if it is an *inner* utterance (i.e., an exercise of one’s capacity for auditory imagination used to create an

¹⁰ And, scientific evidence suggests, through feedforward signals within the brain from motor to sensory areas. See, e.g., Kandel, Schwartz and Jessell (2000).

image of uttering that sentence). After all, the inner utterance is in the form of an imperative, and imperatives, the will, desires, intentions and the like all share the same direction of fit. Even so, an inner utterance of "Get that chocolate!" isn't itself an instance of consciously willing. An inner utterance, like an outer utterance, is one which may be made sincerely or insincerely. If you say "Get that chocolate!" to yourself because you are urging yourself to acquire some piece of chocolate, your inner utterance is sincere and expresses your willed intent for the chocolate. On the other hand, if you say "Get that chocolate!" to yourself because you are trying to think up examples of imperative sentences and that one pops into your head, or because you are recalling making that utterance, then your inner utterance isn't sincere and does not express your will (even if you're willing yourself to get chocolate). In these latter cases the reason for the inner utterance would not be to express to oneself the inner willed action to get chocolate, it would be for another reason, say to think of an example of an imperative. But it may be incidentally true that one actually wants chocolate at that moment. These cases, if at all plausible, make it clear that the will does not cause the inner utterance, even in cases where the utterance is in accord with how we will. there is no necessary connection between the utterance and the will. The possibility of sincere and insincere inner utterances, like the possibility of sincere and insincere outer utterances, shows that there is a gap between the thing that may be sincere (the utterance) and the thing that makes it sincere (your will). Since the sincerity of the utterance can vary, regardless of how we might actually be willing, the inner utterance cannot be the will. But since then the thing that is in consciousness, the utterance, isn't the will my position is still tenable.

Perhaps it is best to characterize what is in consciousness in this case as an expression of the will but not the will itself.

If none of the above elements of consciousness are the actual will then why do people claim that their wills are conscious? There are characteristic feelings—experiences and inner utterances—*caused by* the will. But a feeling caused by the will no more makes for a conscious will than a feeling caused by a brick makes for a conscious brick. Or, to keep to psychological states, a feeling caused by the will no more makes for a conscious will than a feeling caused by one's implicit knowledge of syntax makes for conscious knowledge of syntax. The so-called "conscious will" is revealed by phenomenological inspection to be better described as a will *of which* we are conscious.

It is time to take up some possible objections. I will consider two:

One might hold that the experience of consciously willing something to happen is sufficient for the will to be in consciousness. If the will is something we are conscious of then in some sense the will is in consciousness. Certainly this is a common way of talking about our experiences and the world. And if the will is in consciousness in some sense then there is no separation between consciousness and the will and I must be wrong in claiming that the two are distinct.

The problem here is that there are two ways of talking about consciousness. There is a way of talking about consciousness in which consciousness is like a bucket. Things can either be in or out of the bucket with little room for ambiguity, either something is in the bucket or isn't. Analogously, either something is in consciousness or it isn't. In this sense it is correct to talk about things being in consciousness. But obviously we are not talking about those objects we perceive being, strictly speaking, in consciousness. It is not the objects themselves that are in consciousness but our experiences of them.

This leads to the second way of talking about consciousness. Some people talk about consciousness, not as a space to be filled but as a thing made up of other things, in much the same way that we talk of a wall being made up of bricks. Here we are talking about the constituent parts of consciousness, what consciousness is made of, not what we are conscious of.

But I take the second characterization of consciousness to be more accurate. This is because when we speak of things being in consciousness we never mean that the object itself is in consciousness, it is merely a way of speaking. When one says that the table in front of them is in consciousness they don't mean that the object, the table, is in consciousness. What they mean is that they are experiencing the table and that experience is in consciousness. This is more compatible with the second way of characterizing consciousness because experiences are one of the things we think consciousness is made up of. We think that experiences are at least part, if not all, of

the building blocks of consciousness. So the second way of characterizing consciousness is more accurate.

But if this is right then the will, strictly speaking, is never in consciousness, it is something we have experiences of and those experiences make up consciousness. The will is only in consciousness in some loose, inaccurate sense, although a sense that people commonly use. Nothing about us experiencing the will gets the will into consciousness, in any accurate sense. So since the way of talking about consciousness that gets us to talking about the will being in consciousness is inaccurate it seems as if we have little justification, at least on those grounds alone, for the above claim that the will is actually in consciousness.

Secondly, one might argue that some kind of reductio can be performed on my position: If it is the case that our will isn't a proper part of consciousness then our phenomenology should reflect that. It should feel as if we, our conscious selves, are just along for a ride while our will is moving us around. It *should* feel that way if I'm right but it doesn't, so my position must be wrong somewhere along the line. My response to this has already been hinted at but is worth repeating. Even though the will isn't part of consciousness we are *conscious of* the fact that we are willing ourselves to behave in a certain way. So it is in fact true that consciousness is just along for the ride, but there is no need for it to feel as though we have no control since we are constantly being given information about how we are willing ourselves to behave. A more savvy version of the argument might hold that our feelings of

consciously willing don't feel like perceptions at all, when we feel as though we are willing something to happen it feels as though that feeling exerts some kind of control over things. So, this person might argue, why should we accept your conclusion since that entails that things are otherwise than how we directly perceive them to be?

Here, the empirical evidence becomes important once again. If conscious feelings of willing come before the actual willing then it is impossible to explain Libet's results. If consciousness controls our actions then why does the neuroscience seem to commit us to the fact that the basal ganglia controls our actions and that the basal ganglia isn't a part of the brain of which we are conscious? Much more will be said on this issue in Chapter 5 but suffice it to say that if one wants to hold that the phenomenology of consciously willing must be indicative of how things are then a lot of empirical evidence will either have to be ignored or dismissed. This seems unreasonable.

I conclude that instances of willing and every other propositional attitude having the world-to-mind direction of fit exclusively are never found within consciousness. All will-like states found in consciousness are experiences or exercises of imaginative capacities pertaining either to the will or to the content of the will, but never to just the will itself. Though it was once taken for granted that consciousness is the essence of the mental, though a number of contemporary philosophers have continued to defend this view, and though it appears simple common sense to say that sometimes

our conscious lives include consciously willing as an element, such views appear ultimately untenable.

Chapter 5: Conclusion

In chapter 1 I introduced a problem to be solved. The problem was that, given Libet's results, it seems as if our will precedes our feelings of consciously deciding to act. This was supposed to be a problem for our everyday conception of the will in that we usually think of the will as a part of consciousness, but if the willing comes before the perception of the willing then obviously there is some separation between the two. The feeling of consciously willing oneself to do something isn't the same thing as the actual willed intention. So how do we resolve this problem? One option might be to deny that it exists, to claim that our phenomenology is correct no matter what the science says. Hopefully I have given enough scientific evidence up to this point to make this seem dubious, but just in case I haven't this issue will be dealt with here again. Another option might be to argue that the will and consciousness must be related in some sense and so we have reason to doubt anything that says otherwise. Searle, Strawson, and Descartes can be seen as doing just that, offering arguments to support the thesis that the will and consciousness are intimately linked, although not in response to Libet. Lastly someone might argue that no coherent picture can be given that makes sense of the phenomenology, the empirical evidence and the philosophical arguments in favor of separating the will and consciousness. Since no coherent picture can be drawn there must be something wrong somewhere along the way. In this chapter I hope to show that a coherent picture can be drawn that accounts for all of these things and that, because it can account for everything, it has more explanatory force than any of the other options.

The picture, in its simplest form, is easy to understand and has already been alluded to. Our will is something we are conscious of but not something that is in consciousness. One might think that I have just given up the game, admitted defeat. Didn't I just claim that the will was something we are conscious of which would entail, someone might say, that there is a conscious will? No. I did just claim that we are conscious of our will but this is a very different thing from the will being in consciousness. As I said in Chapter 4, consciousness of the will no more makes for a conscious will than consciousness of a brick makes for a conscious brick. This distinction between us being conscious of something and something being in consciousness is tricky, so let me try and flesh it out a bit.

Anything you can experience can be in consciousness, in the sense that you are conscious of it. But you being conscious of it doesn't make it a part of what consciousness is. Consciousness isn't made up of things like tables, the sun or the number 2, but we can be conscious of all these things and so sometimes they are spoken of as being a part of consciousness. But obviously you don't have a table, the sun or the number two somewhere in your head. What you do have are experiences of these things and it is the experiences of these things that makes up or constitute your consciousness. Think of the will in the same way; we have experiences of our will and so we are conscious of it, but the will does not in any way constitute, or make up your consciousness. Nor does our consciousness make up our will, in the following sense. One might think that if we are conscious of the decisions we make about how to act then we have the right information in order to be able to make decisions in

consciousness. In Chapter 4 I argued there that nothing found in consciousness is the will but if decisions are in consciousness and the will is what decides how to act then why isn't the will in consciousness? Again, I hold that these decisions are only in consciousness in the sense that we are conscious of them. Consider the following: If you are not a dualist (or perhaps a behaviorist) then you think your consciousness is located somewhere between your ears. But when you say there is a table in your consciousness you don't mean that there is a table between your ears. You mean that you are conscious of the table. But being conscious of the table, again, doesn't make the table part of your consciousness it only makes it, in some sense but not an accurate one, in consciousness. The same is true of decisions, there are decisions that we make that we become conscious of but, like the table, that does not entail that the decisions are themselves in consciousness. So speaking properly, there are no decisions in consciousness. But if there are no decisions in consciousness and the will is what makes decisions about how to act then the will isn't in consciousness.

If this is right then it is relatively easy to solve Libet's puzzle. The will is found in the basal ganglia and very shortly after activity there we become conscious of having willed something. Since we are conscious *of* willing it has to do something before we can become conscious of it. This idea is also in accord with all of the empirical evidence presented in chapter 3 and the philosophical arguments in chapter 4. It also explains how it is that we could feel as though the will is a part of consciousness and yet it not be. So it seems as if we have a theory that explains the empirical evidence, the phenomenology and the philosophy.

But there are other contenders. Perhaps one of the theories in Chapter 2 could better account for everything. Perhaps Strawson, Searle, or Descartes has better arguments than I do and so his theory should trump mine. In what follows I hope to show that this isn't the case.

It seems to me that Descartes is simply misguided about the nature of the mind. I think it would be impossible under Cartesian dualism for something to be a product of our minds and yet for us to not be conscious of it. This is obviously false. Since Freud, at least, we have thought there was a conscious and a non-conscious part of our mind. But even without Freud it seems obvious that some of our mental states are not presently in consciousness. I have any number of beliefs in my memory that I am not conscious of right now, if I didn't I could never remember what I believe or there would be no point to such an exercise. So the primary fault with Descartes is that his basic presupposition, which leads him to the idea that perceiving our will and the will are the same thing, is false. Were one not a Cartesian dualist there would be no need for this position.

Searle and Strawson are significantly more difficult to deal with. Both think that experiences make the difference between intentional mental states and non-intentional mental states. Searle holds that experiences in consciousness make the difference between thinking about water and H_2O . Strawson holds that if it were not for experiences we would be functionally identical to thinking machines, machines

that lack intentionality. Since consciousness is where experiences happen, they require consciousness for intentionality. Since the will is an intentional mental state it needs intentionality and, therefore, consciousness. But if consciousness is required for willful actions then I would be wrong, at the very least, in my assertion that the will and consciousness are separable.

Some things can be said about both Searle and Strawson that might make their positions seem less appealing though. Both Searle and Strawson require consciousness for intentionality, they think that the only way for our mental states to have intentional content is for us to have experiences. But this isn't the only view of intentionality. Other philosophers have different views on intentionality. Dennett (1991), Millikan (1995) and Fodor (1975), to name a few, each have views on the intentionality of mental states that do not require consciousness. So although Searle and Strawson might be right about intentionality and consciousness the debate is obviously not settled.

One additional point can be made about Strawson. Strawson holds that the essential nature of desire is to dispose us to pleasure or displeasure upon an instance of that desire being satisfied or frustrated. Since I'm willing to agree with Strawson that pleasure and displeasure are in consciousness (in the sense that they are experiences and experiences make up consciousness) one might think that I should hold that desires are as well. Certainly this seems to be Strawson's position. But this all rests on a category mistake of the following sort: If desires are dispositions to pleasure and

displeasure then they are not themselves pleasure and displeasure, they merely dispose us to those feelings. But then why hold that desires are a part of consciousness? It seems what Strawson really is really committed to is that desires have some kind of effect on consciousness. If he were to continue to hold that desires themselves are a part of consciousness, in the sense that they actually make up consciousness not in the sense that they are something we are conscious of, then it is his burden to explain how that might work. I have no conception whatsoever of what it would be like to have a disposition be a part of my consciousness and I'm willing to bet that no one else does either. This is primarily because when we think about the kinds of things consciousness might be made up of dispositions don't seem like the kinds of things that one could use to build consciousness. So Strawson has confused the product of desire (pleasure or displeasure) with the desire itself (the disposition) and thereby falsely claimed that desires should be a part of consciousness. Of course one may hold that some aspect of the desire, the conscious part is in consciousness and that there is another aspect that is not. So, on this view, maybe the disposition isn't in consciousness but the resulting pleasure or displeasure is. One might then claim that, in the same way that a boat can be in water without being submerged, the desire is in consciousness. The primary problem with this view is that pleasure and displeasure are not solely the result of desire satisfaction, as seems obvious from the fact that unexpected things can often cause us pleasure, even in cases where we don't. So the resulting phenomenology is separable from the desire satisfaction. But if this is the case then why think that the phenomenology and the disposition are both one and the same thing? Why not hold that one is the cause of the other, which seems much

more plausible? Though if the disposition is the cause and the phenomenology the effect they are obviously not the same thing. It doesn't seem plausible to hold that both are aspects of the same desire.

Another problem for both Strawson and Searle is that I have introduced what I take to be compelling scientific evidence that the will is a separate entity from consciousness. Neither Strawson nor Searle offer any empirical evidence to back up their claims. One might think that any theory with empirical support should trump another theory lacking it but even if one doesn't I hope I have at least made it seem as though Strawson and Searle have some things to answer for before we accept their theories wholesale.

Wegner also has a theory that would solve Libet's puzzle. He thinks, as I do, that the empirical evidence points to there being a separation between the will and consciousness. But he has a different way of explaining how it is we come to feel as though we are in control of our actions. Before saying why I don't think one should adopt Wegner's position, even though it would explain everything that I'm trying to explain, perhaps I should say what Wegner's position is.

In his book *The Illusion of Conscious Will* Wegner holds that our feelings of being in control of our actions, though illusory, are the result of the fact that our actions so frequently accord with how we intend to act. When we have intentions and our bodies act in the right way to carry out those intentions we get a conscious feeling of being

the author of those actions. This, according to Wegner, is where the illusion of conscious will comes from.

Conscious will is the mind's compass. As we have seen, the experience of consciously willing action occurs as the result of an interpretive system, a course sensing mechanism that examines the relationship between our thoughts and actions and responds with 'I willed this' when the two correspond appropriately. (Wegner, 2002, 317)

Why, if this experience of will isn't the cause of action, would we ever go to the trouble of having it? What good is an epiphenomenon? The answer becomes apparent when we appreciate conscious will as a feeling that organizes and informs our understanding of our agency. Conscious will is a signal with many of the qualities of an emotion, one that reverberates through the mind and body when we sense having authored an action. The idea that the conscious will is an *emotion of authorship* moves beyond the standard way in which people have been thinking about free will and determinism, and presses towards a useful new perspective. (Wegner, 2002, 318)

Wegner and I are obviously in agreement over some things. We both agree that our feelings of conscious will don't have the causal power normally associated with them. They have no real causal power over our actions. But where we differ is Wegner's

view on what these feelings of conscious will consist in. Wegner believes that the feeling of consciously willing causes a feeling of agency (or control) and this feeling (or emotion) is the mind's compass (whatever that means). Here Wegner is claiming that the reason for our feeling of agency is that we feel we consciously will ourselves to act and even though this feeling is epiphenomenal it very often accords with how we actually act. This then makes us feel like we are actual agents. I disagree with Wegner about that this. Here's why.

So we think about performing some action, and then once we actually perform it we receive a feeling of authorship that causes us to believe that we consciously willed the action to happen. But what of those who can't act, or who can't act in the way that they intend? Here I mean people who are either paralyzed or have some sort of Parkinsonian disorder. Because their bodies either cannot act, or cannot act in the way they intend them, under Wegner's theory it seems as if these people should have no feeling of authorship and, therefore, no feeling of consciously willing action since this feeling of authorship just is what gives us the illusion of conscious will. I take this to be highly implausible. People who are paralyzed or have neurological disorders can consciously experience themselves as willing their bodies to do something in just the same way that normal people can, their bodies just don't respond in the same way. So if these people feel as though they can consciously will themselves to do something then there must be something wrong with Wegner's positive theory. It cannot be the case that our feelings of conscious will are a product of our actions according with

our thoughts, although this surely plays some role in why we feel we are in conscious control of our actions.

But, one might think, Wegner's theory explains how it is that we come to have a feeling of agency and my theory doesn't, or hasn't. So Wegner's theory has a major virtue that my theory lacks. Let me explain how I think it is that this feeling of agency comes about. I think, to a certain extent, that Wegner is right. When we will ourselves to do something it comes from us and because it comes from us we get a feeling that it was the product of something we initiated. So we feel as though the action is ours. The problem with Wegner's theory is that he holds that this is the only way to achieve this feeling of agency. On my view there is an additional factor that can contribute to this feeling of agency. Under my view we are sometimes conscious of how we are willing ourselves to act and sometimes not. But in those situations where we are conscious of the willing it feels as though the will is ours because, consciously or not, we realize that sometimes we get ourselves to do things without being conscious of what we are willing. While sitting here typing I am not conscious of willing every key stroke. If I were my typing speed would significantly decrease. So I let the non-conscious part of me take over and I worry about what I am going to say¹¹. But if this is true then I can either be conscious of what my will is doing or not be. Perhaps, and this is conjecture, when I am conscious of my will it feels like it is mine in a sense that it does not when I am paying attention to something else. But if there are times when it feels as though I am doing the willing and times when it does not then

perhaps there are times when it feels as though the will is mine and times where it does not. Perhaps the feeling of agency arises, in part, from the fact that sometimes we feel in possession of our will and sometimes we don't. Agency then results from the feeling that the will is being consciously guided in some instances as opposed to instances where it is not.

I don't plan to develop a full theory of agency here but hopefully I have given some hints that might fill in the gaps of Wegner's theory. If it is the difference between consciously and non-consciously willing that leads to agency, or if it's the fact that choice seems like it has to be a product of consciousness then one can explain why paralyzed of Parkinsonian people still have a feeling of agency in a way that Wegner's theory cannot.

Along related lines, perhaps another reason for our feelings of agency has to do with choice. In some sense we normally think of choice as being a conscious process. If I didn't consciously choose to act this way then it isn't my action, or so popular opinion goes. Although it's fairly simple to speak of non-conscious intentions occurring in day to day life the idea of non-conscious choice is harder to swallow. Perhaps the feeling of agency results, in part, from the fact that there is a feeling that the choice has to be ours. .

¹¹ Of course, typing isn't like alien hand syndrome. I am still aware of what is being done by my body and why it is being done. I haven't lost track of why I am doing things, I am just not currently feeling as though I am consciously willing something to happen.

So the solution to Libet's puzzle is that our will is not a part of consciousness but something that we become conscious of. Empirical evidence has been presented to back up this claim. Philosophical arguments have been made to support this conclusion. Although I haven't annihilated the competition I hope that all of these things working together makes my view seem appealing. In the end I think the will is a separate and distinct entity from consciousness, for all of the reasons presented. I hope I have motivated the reader to consider this as an option as well.

References

- Anscombe, G. E. 1957. *Intention*. Oxford: Blackwell.
- Churchland, P.M. 1995. *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.
- Dennett, D.C. 1991. *Consciousness Explained* Boston: Little, Brown & Company
- Descartes, R. 1988. "The Passions of the Soul," in *Descartes: Selected Philosophical Writings*. J. Cottingham, R. Stoothoff and D. Murdoch (trans.). New York, NY: Cambridge University Press.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Fodor, J. 1974 *Language of Thought* Thomas Y. Crowell Co.
- Frankfurt, H.G. 1988. *The importance of what we care about: Philosophical essays*. Cambridge University Press : NY, 1988
- Gomes, G., 2002, Problems in the Timing of Conscious Experience. *Consciousness and Cognition*, Vol. 11, 191-197
- Humphries, M.D. and Gurney, K.N. 2002. The role of intra-thalamic and thalamocortical circuits in action selection. *Network* vol. 13, 131-56.
- Jacob, P. 1995 Consciousness, Intentionality and Function. What is the right order of explanation? *Philosophy and Phenomenological Research*, vol. LV, 195-200
- Jeannerod, M. 1997 *The Cognitive Neuroscience of Action*. Blackwell. Malden, MA, US

- Kandel, E.R., Schwartz, J.H and Jessell, T.M. 2000. *Principles of Neural Science*(4th ed.) McGraw-Hill.
- Libet, B. Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *Behavioral and Brain Sciences* vol. 8 529-566.
- MacKay, D.M. 1985, Do we control our brains. *Behavioral and Brain Sciences* Vol. 8(4): p. 546
- Millikan, R.G. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass. MIT Press.
- Mink, J.W. 2002. The Basal Ganglia and Involuntary Movements. *Archives of Neurology* (60) 1365-1368
- Mink, J.W. 1996. The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology* vol. 50, 381-425
- Naatanen, R.J. 1985, Brain physiology and the unconscious initiation of action., *Behavioral and Brain Sciences* Vol. 8(4): p. 549
- Ringo, J.L. 1985, Timing volition: Questions of what and when about W. *Behavioral and Brain Sciences* Vol. 8(4): 550-551
- Roth, G. 2003. "The interaction of cortex and basal ganglia in the control of voluntary actions" in *Voluntary action: Brains, minds, and sociality*. Maasen, Sabine, Prinz, Wolfgang (Ed) London, Oxford University Press.
- Schroeder, T. 2004 *Three Faces of Desire*. Oxford University Press.
- Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Searle, J. 2002. *Rationality in Action*. Cambridge, MA: MIT Press.

- Strawson, G. 1994. *Mental Reality*. Cambridge, MA: MIT Press.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Van Gulick, R. 1995. "Why the connection argument doesn't work." *Philosophy and Phenomenological Research* 55(1): 201-207.
- Wegner, D. 2002 *The Illusion of Conscious Will*. Cambridge, MA: MIT press.
- Wichmann, T. and DeLong, M.R. 1996. Functional and pathophysiological models of the basal ganglia. *Current Opinion in Neurobiology* vol.6 751-8
- Wood, C.C. 1985, Pardon your dualism is showing. *Behavioral and Brain Sciences* Vol. 8(4): 557-558