

https://dspace.org/

Whether we have free-will and whether it matters

Ostrowick, John Montague

2007-03-01T12:33:53Z
http://hdl.handle.net/10539/2182

Downloaded from a DSpace repository. https://dspace.org/

WHETHER WE HAVE FREE-WILL, AND WHETHER IT MATTERS

John Montague Ostrowick

A dissertation submitted to the Faculty of Humanities, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Arts in Philosophy.

Johannesburg, 2006

ABSTRACT

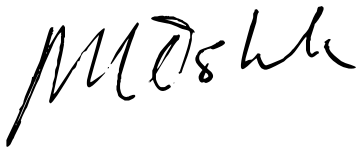
There is a concern that causal determinism might render free-will impossible. I compare some different perspectives, namely *Compatibilism*, *Incompatibilism*, *Libertarianism*, and *Hard Determinism*, and conclude that Hard Determinism is correct—we lack free-will. To further bolster the case, I consider the work of Libet, who has found neuropsychological evidence that our brains non-consciously cause our actions, prior to our being aware of it. Thus we are also not choosing consciously. I then consider Dennett's work on the role of the conscious self. I defend his model—of a fragmented self—which could not cause our actions. Finally I argue that many things that free-will purportedly provides, eg., justification for the penal system and reactive attitudes, can be reconstructed without free-will. I then end with some speculations about why people still want free-will.[1]

(**Keywords**: *Compatibilism*, *Incompatibilism*, *Libertarianism*, *Hard Determinism, Free-will, the Self, Libet, Dennett, Reactive Attitudes, Penal system*)

[1] I will omit the theological debate about free-will for reasons of its relevance. This is the debate about how it is that an omnipotent, omniscient, omnibenevolent God could allow us to do wrong, through exercise of our free-will, and how it is that we could even have free-will at all if all the events in the world were due to His will.

DECLARATION

I declare that this dissertation is my own unaided work. It is submitted for the degree of Master of Arts in Philosophy (by dissertation) in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination in any other university.



John Montague Ostrowick

31st day of March, 2006.

ACKNOWLEDGEMENTS

I wish to thank first, my supervisor, Prof. Mark Leon, for his patience, insight, and precision in reviewing my work and in helping me to see my many errors, as well as Prof. Daniel Dennett, to whom I owe my initial interest in this topic, and without whom approximately half of the novel material in this dissertation would not exist. It was Dennett's fascinating book, *Consciousness Explained*, which more than any other inspired me to write this dissertation, and in which I first saw mention of Libet's important work. I would also like to express my gratitude to my friends and family for their support and faith in me. In particular I wish to thank Dr. Brett Bowman for keeping me inspired and for the endless conversations we have had on this topic. I also wish to express my gratitude to the department that I have the good fortune of being employed by, viz., the University's Computer Science School—who have allowed me large spans of leave to attend to my writing and who have encouraged me to continue to study in a field largely unrelated to their own. In particular, I wish to thank Prof. Conrad Mueller for allowing me this, and for helping me to obtain financial support from the Science Faculty to pay for my studies over the years. Finally, I wish to thank all the people who drew my attention to pieces of writing that have helped me produce a more interesting work—both persons external and internal to the University. All other materials used in this work have been referenced in the references section at the end of the document.

—J.

CONTENTS

		Page
1	**Framing the problem**	**1**
1.1	What we stand to lose	1
1.2	Definition of terms and concepts	2
2	**Compatibilism**	**5**
2.1	Introduction	5
2.2	Characteristic Compatibilist Arguments	6
2.3	"Reasons" and "Tracking" Models of Freedom	17
3	**Incompatibilism**	**27**
3.1	The Defence of Incompatibilism	27
3.2	Libertarianism	29
3.3	Hard determinism	46
4	**The Timing Experiments of Libet and Grey Walter**	**50**
4.1	Introduction	50
4.2	The Experiments	51
4.3	Criticisms and Basic Problems	56
4.4	Libet's "Veto"	68
4.5	Summary and Conclusion	73
5	**The Existence, Nature, and Function of the Self**	**75**
5.1	What is a self, and why is it needed for moral responsibility?	75
5.2	Models of the Self—Introduction	79
5.3	The Cartesian Model of the Self	80
5.4	Constructivist Models of the Self	87
5.5	The Skeptical Model of the Self	91
6	**Whether Free-Will Matters**	**98**
6.1	Whether we need free-will and whether it matters	98
6.2	Why we believe in free-will—Some speculations	118

Appendix

References

LIST OF FIGURES

		Page
4.1	Libet's Findings—1	52
4.2	What we believe about choice	54
4.3	Libet's Findings—2	54
4.4	The real concern	54
4.5	What seems to be happening	56
4.6	Libet's Graphs of Action and Veto	68
4.7	A possible solution	73

CHAPTER 1

Framing the problem

> We have certain notions of what freedom is. This dissertation will consider the notions we have and question their legitimacy. But before we can discuss the notion of "free-will" or "free choice", we need to establish the meanings of the terms we will use as well as explicate the nature of the problem.

Section 1—What we stand to lose

Suppose that we are part of a mechanistic world, in which we, like any other physical entity, are capable of interacting causally: We can be affected by causes—experience the pressure of some causal event—and we can also initiate causal events, creating subsequent effects. Suppose, now, that the whole of the world were governed by causation to such an extent that given certain causal antecedents, certain effects would be completely predictable, every time. Many physicists believe this to be a fact about the world, that every event is merely an effect of prior causes, without which the event would not occur. Now suppose that our acts were also mere events, necessitated in this way by antecedent causes. The view that this *is* the case, is called "determinism". Supposing then that determinism were true, would it still be meaningful to say that it was really *us* choosing to do what we do? Or would it not perhaps rather be the case that we are merely cogs in a universal machine, just doing what we "have to" because of prior causes?

> "If determinism is true, then our acts are the consequences of the laws of nature and events in the past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our present acts) are not up to us." (Van Inwagen, *Preface*, i, 1983).

But it may be even worse than that. If everything we did was predetermined by factors beyond our control, it may also be true that we would not be entitled to claim to deserve anything either—not the credit for writing large dissertations, nor the punishment for driving too fast or killing six million people. But surely we could not bear to live in such a world?

> "I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other mens' acts of will. I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me... from [the] outside... " (Isaiah Berlin, in Double, p12).

We *do* have the impression that we can choose what we do. And we do feel entitled to being praised for achievement, punished for wrongdoing, etc. Hence the notion of free-will seems to have a problem. The problem is this: Does determinism threaten our freedom—as we think it

may? And if determinism threatened our freedom, would it matter? In other words, could we still be justifiably rewarded, punished, etc., even if it turned out to be true that we exist in a deterministic context? My job in this dissertation will be to answer these two questions.

Section 2—Definition of terms and concepts

Before we can proceed with our considerations about the notion of freedom, it is important to provide some definitions of the various terms that will be used. The problem with defining the concept of freedom from the outset is that its commonplace definition(s) are dubious. Indeed, it may be *because* the commonplace understandings are problematic that "the problem of free-will" exists at all. As we shall see, on investigating this matter, we will have to change or clarify our understanding of the concept of freedom. So, as a rough guide in the meantime, we will use the definitions that follow.

a. Freedom and Free-will

To be *free*, according to the Oxford Dictionary, is to be:

> "*adj* **1** not in bondage to or under the control of another [person or agency] ... **3** **a** unrestricted, unimpeded; not restrained or fixed. **b** at liberty; not confined or imprisoned. ... **d** unrestrained as to action; independent ... **4** **a** not subject to; exempt from... **b** not containing or subject to a specified ... thing. **5** able or permitted to take a specified action (you are free to choose). **6** unconstrained. ... **8** **c** clear of obstructions. **9** spontaneous, unforced. ... **17** *Physics* **a** not modified by an external force ... *adv* **3** ... **free agent** a person with freedom of action"

The Oxford Dictionary defines *free-will* (under the entry "free") as:

> "*v.intr*. **free-will** **1** the power of acting without the constraint of necessity or fate. **2** the ability to act at one's own discretion"

and defines *freedom* as:

> "*n*. **1** The condition of being free or unrestricted. **3** the power of self-determination; independence of fate or necessity. **4** the state of being free to action. ... **6** the condition of being exempt from or not subject to ..."

In defining "free-will", it is important to distinguish *willing* from *acting*. Frankfurt also makes this distinction in Watson (1982 p90): Freedom of *will* is a separate question about an agent's mental states. Having "freedom" means to not be restricted regarding possible actions. Having "free-will" refers rather to one's choices or desires. This paper focuses on the psychological questions and draws inferences from that, to the possibility of freedom of action. In other words,

we could not have freedom of action without freedom of will, since we assume the *will* causes our actions. Philosophers express a wide variety of definitions of "freedom", as we shall see, and differ quite strongly amongst themselves on the matter.

b. Determinism. This is the view that every event is causally necessitated (Glover, p454), ie., given the prior states of affairs, the outcome could not be otherwise.[2]

> "Determinism is the view that there is only at one time a single physically possible future world." (Van Inwagen, 1983, in Double, p18).

> "Determinism is the view that for all events there exists a sufficient causal condition that [alone] results in that event." (Chisholm, R., 1976, in Double, p18).

c. Incompatibilism is the view that *if freedom is the case, determinism is not the case,* and *vice versa*. Freedom is incompatible with causal determinism. There are two subvarieties of this view. The first is called "libertarianism". The other is "hard determinism".

d. Hard determinism. This is the view that everything, including the actions of mankind, are causally necessitated *and* that incompatibilism is the case. In other words, *because* we are determined, we are not free.

e. Libertarianism. This is a form of incompatibilism, because it affirms that freedom and determinism are incompatible (see eg., Watson, 1987, p164, Van Inwagen (1983), Kane). Libertarians *deny* determinism and argue that we are inherently free.[3] The libertarian denies that our actions are a product of causal law (Glover, p454).

f. Compatibilism or Soft Determinism. This is the view that we could be free even if determinism is true. Compatibilists hold that free-will is compatible with determinism. Some compatibilists argue that determinism is in fact required for us to have free-will. For this reason the doctrine is sometimes called "soft determinism" (William James in Glover, p455).

g. Hard incompatibilism. This is the view that free-will is incompatible with not only determinism *but also with* partially random causes, *or* with fully random causes or indeterminism (Pereboom, p127). Pereboom says (p89):

> "an action is free in the sense required for moral responsibility only if the decision to perform it is not an alien-deterministic event, nor a truly random event, nor a partially random event".

---

[2] Determinism is not the same thing as fatalism. A fatalist sees an agent as unable to make a difference to what happens in his life; what happens to him will happen *come what may*. A determinist sees an agent's life rather as a product of his circumstances and his responses to those circumstances. For a determinist, the agent's will or actions do make a difference to what happens. We will not discuss fatalism further in this paper.

[3] See eg., Sartre, p26 et seq., Grossman, p251 et seq.

Since all actions are either random, partially random, or determined, the hard incompatibilist will argue, there are no free actions.

h. Skepticism. There are at least two forms of skepticism about free-will. The first, expounded in Double, is that our intuitions about free-will are so incompatible as to make the term itself impossible to fully capture in a single consistent perspective without leaving out some substantial aspect of it. Double's view is not discussed in detail in this dissertation as the bulk of the material presented here depends on the concept of free-will being coherent. The other form of skepticism we will cover in some detail, however, and it involves questions about the prerequisites for having free-will and whether these pre-requisites are actual in any way. More on this will appear in the relevant section.

CHAPTER 2
Compatibilism

Compatibilism is the view that determinism is compatible with our having free-will. Compatibilists argue that what is important for free-will is that an agent have a desire, and the agent gets what she wants because of the desire. What is relevant, for moral responsibility, is whether an agent causes an event and the event she causes is the one that she desired to have happen. Problems for compatibilism, and the issue of whether persons ever have alternative possibilities, are also discussed.

Section 1—Introduction

The question of freedom of the will, for compatibilists, is a matter of the involvement of an agent's "will" in the determination of an action. Compatibilists are concerned with what makes an action voluntary. If an act is voluntary, for the compatibilists, it is free. Let us take a few quotations in order to characterise this perspective. Locke[4] puts it this way:

> "8. ... so far as a man has power to think... [or] to move... according to the preference or direction of his own mind, so far is a man free. ...".

Hume says that a person is free when he is able to act according to the determinations of his will (§73, *Of Liberty and Necessity*). Watson, a modern compatibilist, says: an agent is free if her actions track her will (Watson, 1987, p145). Let us take, for example, the case of someone who has her money snatched, as opposed to when she hands it over voluntarily. In the case where the person hands her money over voluntarily, a compatibilist argues that she has free choice, because she has a certain desire to hand over the money, and she does in fact hand it over. But in the case where the money is snatched, *her will does not intervene at all*.[5] And it is in *this* case—where the will is *absent*—that we regard the person's yielding of her money as unfree. The difference, then, as to whether some action is free or not, is whether an agent's *will* is involved in the action or not.

Compatibilists also do not deny causal determinism. In fact, some compatibilists argue that determinism is *required* for freedom. For consider: if our actions were not determined by our desires (our "will"), then the actions would not be truly our own. So for us to be free, some form of determinism may be a prerequisite.

It may not be obvious at this point, however, that the compatibilists are right. We therefore need to consider their arguments in more detail.

[4] Locke, J. (1690), Chapter XXI, Of Power

[5] This discussion is based on an argument in Frankfurt (1969), pp833–4. See also Stampe, 1992.

Section 2—Characteristic Compatibilist Arguments

Compatibilists generally have a number of characteristic arguments or argument forms, that they utilise in presenting their perspective, and a number of issues that they explore.

> a. The first of the matters that some compatibilists deal with, is their approach to "alternative possibilities"; the requirement that the agent, in acting, must be able to *do otherwise* or *choose otherwise*, in order for that action to be "free".
>
> b. The second feature of compatibilist argument typically involves identifying and describing a mechanism for freedom. Compatibilists want to know in virtue of which mental state or property we are said to have freedom of will. How, in other words, could free-will be instantiated? For once we can distinguish free from unfree actions, we can then attribute the free agent with moral responsibility, if the agent has such a mental state or property which imbues that agent with free-will.

I will discuss these in turn. "Tracking" models, which represent an improvement on the traditional models, warrant a section of their own, and are discussed in section 3.

Section 2.a.1—Arguments for Alternative Possibilities

Some writers feel that an important feature of being free is being "able to do otherwise"—that it is not just "decisions going through one's will" which is important. Consider the typical case of someone expressing regret for his actions. He will say things like "I should not have done that, I should have done something else", or "I could have done it better", etc. We will look at the possible meanings of such phrases. After that we will consider, briefly, whether free-will even requires the notion of alternative possibilities.[6]

Frankfurt (1969, p829), discusses this issue. There are two ways we can interpret "alternative possibilities": the *conditional* and the *unconditional* interpretations. (These terms are also used in *Van Inwagen*, 1989, p403 et seq., *Leon,* 2002a).

i. The **unconditional analysis** of "could have done otherwise" (UCA) amounts to a suggestion that *to be free* means *to be able to act differently regardless of the situation being the same*. Consider the words of a person expressing regret: "If I were to live my life over, I would have done it differently", "I should have done something else", etc. All these phrases can be interpreted, on the unconditional analysis, to mean something like: "If my life were to be re-instantiated with all things being held the same, I would have acted differently". Superficially, it means that we could be in an identical situation and yet do something different. But that is impossible—if you accept the reality of determinism. For consider: the agent's wishes and

[6] Some compatibilists (eg., Frankfurt) argue that agents could be free even if they could not have done otherwise.

desires were brought about through a deterministic causal sequence. If she were to live her life out again (we presume), she would have the exact same set of desires. And if she had the same set of desires, we would expect that she would act on those desires in the same way, because her desires cause her actions. Leon (2002a, p426 et seq.) puts it this way: Suppose we imagine a parallel world to ours in which we have a Doppelgänger—a completely identical twin who has the same mental states as ours. The *un*conditional interpretation of "could have done otherwise" would require that our Doppelgänger, to be free, would have to be able to *do something different* to what we do. But now suppose that when *we* act, we act for our best reasons. If our Doppelgänger acts differently to us, then she will be acting against her best reasons. The Doppelgänger would, in fact, be acting irrationally (Leon, 2002a, p429). Yet we don't think that an insane person, or a person acting against her best reasons, or a person acting against what she wants, is acting freely. Clearly if the agent acts against her best, strongest or most valued desires, she is not free. "[The] decision process would be an exercise in anarchy" (Double, p197). If an agent cites reasons for acting, then her decision process *could* only explain one choice (*ibid*., p205). Therefore, any choice that was spontaneous in the above sense (lacking causation by reasons) would be irrational or capricious. The unconditional analysis of alternatives thus fails to preserve rationality and control (Double, p194, also in Ginet, 207); and it also fails to provide a satisfactory explanation of what it means, to say that one is free if one "could have done otherwise".

ii. On the other hand, the **conditional analysis** (CA) says this: ***if*** *the will of an agent is different in a situation, then the act will be different*. In other words, *if* the agent had willed to do otherwise, then the agent would have done otherwise. This is called the "conditional" analysis of alternatives, because choosing otherwise, in these circumstances, is conditional on the agent's desires being different. Only if his desires had been different would his act be different. This is compatible with the idea of determinism, because it is not denying that the "*will*" might still be the same. This certainly makes sense of the notion of "being able to do otherwise". For consider: If I say, "I could have done better", it means that at the time I took the choice, it was open to me to perform the action better. And if my will had been different at the time, I would have performed the act better. The conditional interpretation allows this description. It is a form of compatibilism because (a) what counts is that I act according to my will and (b) determinism is not denied.[7] Thus compatibilism accounts for our intuitions about alternative possibilities.

Why do some compatibilists think that alternative possibilities are important for freedom or moral responsibility? Well, the answer turns on whether an agent has a say in what he does. If an agent *could have done otherwise*, then he would have had a choice about what he did. Without alternative possibilities, there will likely be no choice in the real sense of choosing between two or more options. If at the moment of choice, an agent could have done otherwise if he had so chosen (Leon, 2002a, p430), then the agent would have had free-will. But we would need to ask whether the necessary conditions for choosing otherwise were up to the agent or not. The example Leon gives is of a drowning child: An agent is responsible for saving or not saving

[7] see also Van Inwagen, 1983, pp108-9

the child depending on whether or not the agent knew how to swim, realised the child was in danger, etc., and then acted accordingly. Only if the relevant necessary conditions for an agent's actions were not available, would we exempt him from moral responsibility for his action.

This discussion of alternative possibilities seems to be aimed at establishing (a) what alternative possibilities would be, and (b) what "freedom" would mean given a background assumption of determinism. On the best reading (the conditional interpretation), alternative possibilities would mean just this: That one had an alternative possibility under the condition where, had one's will been different, one's act would have been different. This is compatible with determinism, because it is not saying whether one's will was in fact different, or whether the history of the universe would have let it be different, etc. It is merely characterising what alternatives depend on. On the question of freedom, then, an agent is said to be free in the case where she has alternative possibilities. If this were correct, then, freedom would be the circumstances under which an agent's actions tracked her *will*. Free actions are reasons-tracking actions.[8] The point of the conditional interpretation is that actions are free if they track the will, and the will is free if it tracks the reasons that there are for doing something. I am free insofar as I do whatsoever I will, and if my will had been different, I would be free insofar as I did that other different thing. The conditional interpretation of alternative possibilities is not, however, trying to establish that the agent's *will* would in fact ever be different to what it is, or that in some alternative possible world[9] that it would be different to what it in fact is.

The conditional interpretation is also characterising choice using the same model as science characterises other law-like relationships. Compare the structure of these two descriptions: (a) *If I had struck the match, it would have lit*, and (b) *If I had willed otherwise, I would have done otherwise*. They have the same structure, and neither description (a) or (b) rules out either determinism, or the possibility that the counterfactual was not fulfilled (ie., the match was not lit and I did not will differently). Yet both counterfactuals are clearly *true*.[10]

Let us try to apply this model. What is the difference between a drug addict and a normal person who chooses to take in some substance? The answer: A drug addict does what he does, *come what may*—the drug is always the cause of what he does in relation to taking it. A normal person, however, does what he does because of his will; he takes in some substance because of factors other than addiction (viz., choice). The conditional analysis says that if a person is free, and his will is otherwise, he will do otherwise. So a normal person, choosing between taking in a substance or not, can choose either according to his will. A drug addict however, regardless of his will, always will take the substance. Thus he does not fulfil the requirements for the conditional analysis because his will has no effect: His will is otherwise (he doesn't want to take

[8] Where "reasons", means, loosely, our beliefs and desires—cf. Davidson.

[9] "The basic idea is that a possible world is a *way things could have been*; it is a *state of affairs* of some kind" (Plantinga in Peterson, M., Hasker, W., Reichenbach, B., Basinger, D. (1996), pp 270-7).

[10] Leon, 2002b.

it), but he takes it anyway (because of the addiction). Thus *because* his actions do not follow his will, a drug addict is not free. The conditional analysis thus successfully draws the distinction between free and unfree actions. However, we know that an agent's beliefs and desires will be the same if the truth and evidence remains the same (*ceteris paribus*), and we acknowledge that choices would be different if the beliefs and desires had been different. And we want our beliefs and desires to be different if the evidence or truth is different. But if the evidence or truth is held constant (*ceteris paribus*), then we don't want our beliefs or desires to be any different. Thus, we only want our desires to be different if the circumstances are different—the conditional analysis. What we want is for our beliefs to be properly formed (determined by the truth and evidence), and for our desires to be properly formed, so that they follow or track our real needs and true interests, so that if our needs or interests are in fact different, then our desires would be different. It's not just our actions which are conditional on our choices, it is that our choices are conditional on our reasons, and our reasons (beliefs and desires) are conditional on the evidence and our needs and interests, respectively (Leon, 2002b).

Section 2.a.2—Alternative possibilities—Critical Discussion

There are however a few things about the concept of alternative possibilities which are worrisome.

i. The first of the concerns is this:

> "If determinism is true, then clearly, in some sense, there are no alternative possibilities." [*sic*]. "... Relative to the laws of nature and antecedent conditions, it is not possible that one does anything but what one does." Watson (1987), p154.[11]

Neither Watson (himself a compatibilist), nor I, are arguing that there is no meaningful way in which we can talk about alternative possibilities. Rather, Watson is here expressing the same worrying intuition I have about the relevance of the causal past. We *can* characterise freedom in terms of "actions following the will", and it is reasonable to say we are free if our actions *do* follow our will (unlike the drug addict). And we can accept that the counterfactual expressed by the conditional analysis is always true for a "normal" person, and that a person for whom that counterfactual was true, would be "free" in that her actions would follow her desires. But is this "freedom" enough for moral responsibility? When we talk of alternative possibilities (conventionally), we are talking of a choosing *process*, and that choosing process involves two or more future paths. We punish or reward on the assumption that the path we choose or desire is *up to us*, not because we *did* something but because it was (theoretically) up to us *whether* we wanted to do that thing. The problem being suggested here is that ultimately, what we do is a result of what we desire, and it is not obvious that what it is that we desire is up to us. I believe it

[11] This is not meant to disprove compatibilist freedom; Watson is a compatibilist. But it is a direct quotation.

is this intuition which lies at the heart of the incompatibilist's worries about free-will[12]. Glover puts it this way: *"I can do what I want, but can I want what I want?"* (Glover, p458). The conditional analysis of alternative possibilities answers this by asserting that if our needs had been different, what we wanted would have been different. But Watson's point here is just that the laws of nature, and the antecedent causal past, ensure that we only choose one path, every time. "An action is free in the sense required for moral responsibility only if it is not produced by a determinisitic process that traces back to causal factors beyond the agent's control." (Pereboom, p3).

ii. My second concern with the discussion of alternative possibilities (and not a very telling one, admittedly), is that contrary to what the compatibilists may argue, I have a suspicion that people who say "I should have done otherwise", are in fact referring to the *un*conditional analysis. I believe that people use these counterfactual phrases in the implausible, impossible sense, and that *that is why* I think that "alternative possibilities" shouldn't be a requirement for free-will (or that they render free-will implausible). I know that until I was convinced of the error of this idea, that I certainly held alternative possibilities to refer to the unconditional interpretation. That the unconditional interpretation is flawed escapes many who hear it or use it. Expressions like "If only I had done it differently" seem to mean that the person believes it would have been possible to have done something differently under those same conditions. And people persist in thinking that way. I believe, however, that these expressions—where people say "If only I had done otherwise"—are really expressions of regret. But what we've done is really not something we can do anything about. As Dennett says:

> "Suppose I find I have done something dreadful. *Who cares* whether, in exactly the same circumstances and state of mind I found myself, I could have done something else? I didn't do something else, and it's too late to undo what I did. But when I go to interpret what I did, [what is important is] what do I learn about myself?" (Dennett, 1984, pp142-3).

iii. My third concern is this. Recall that at the beginning of this section we mentioned that some compatibilists do not think that we need a concept of alternative possibilities in order to allow for the reality of freedom. Dennett (2002) may be an example. He argues that we can be morally responsible for events even if we could not do otherwise. In causing a moral event, we affirm our characters or our desires—and that's what is important in our being responsible for something. Dennett refers us to Luther, who in breaking from the Catholic Church, said "Here I stand, I can do no other". Dennett says, even if Luther *literally* meant that he could "do no other", it does not mean he was not freely choosing to do what he was doing, or that he didn't want it to be that way. This seems to suggest that alternative possibilities are not required for free choice.

In *The Metaphysics of Free-will*, pp149-153, John Martin Fischer presents an argument, (originating in Frankfurt) which might be used to avoid the question of the origin of an agent's mental states. Imagine a nefarious neurosurgeon who implants a device in an agent's brain such

[12] See also Double, p191 and Chisholm, pp50-51 for a related argument.

that if the agent wills to do anything the surgeon does not desire, the agent's will shall be quelled, and replaced by the surgeon's desire. If the agent wills in accordance with the surgeon's desire, then the intervention will not take place. Fischer argues, and I think convincingly, that the agent is *only* responsible for his actions when those actions are *not* interfered with by the surgeon's device (p152). But this means that *despite* the fact that the agent does not have alternative possibilities (no freedom as such), he is *still* morally responsible at least some of the time. The thought experiment, I believe, is meant to show that determinism is compatible with moral responsibility, and again, that alternative possibilities aren't strictly required for freedom (Leon, Footnote 5, 2002a, p426).

The conclusion, briefly then, is that we might not need a concept of alternative possibilities; as long as an agent can get what she wants, we regard her as "free". Now that we have looked at what compatibilists propose as the preconditions of freedom, let us look at how these writers provide a mechanism for freedom of will, given that we exist in a deterministic universe.

Section 2.b.1—The Nature of Free Choice

Typically, compatibilist writers will privilege some mental entity or system as being "functionally equivalent" to an agent (Velleman, p123). For an agent to be responsible, the agent has to be *in her actions*. So to identify *when* an agent is in her actions, compatibilists postulate the existence of certain special mental systems. Compatibilists characterise free action as action which is internally motivated or which originates in some feature of the agent. The idea of the privileged mental entity or system, then, is to identify which mental system could be sufficiently original or particular to each person, to count as the cause of her actions—the power behind the throne of the *will*. These systems, the idea goes, constitute or accord with the will of the agent, and in virtue of the agent's *will* being constituted by that system, or being in accord with that system, we regard the agent as free. Let's look at a few samples of this argument style.

Frankfurt [13]

Persons, says Frankfurt, are special in that they are the only creatures capable of having second-order desires. These are desires about the content of our desires: "Men may also want to have (or not to have) certain desires and motives." (pp82–6). Frankfurt defines a "wanton" as someone who lacks second-order desires; someone who doesn't care about what it is he desires (p87). We generally are not wanton; we care about *what* it is that we want. A person who cares about what it is that she wants, is a person who has second-order desires: desires about what her effective desires should be. If a person has second-order desires, and the person's actions accord with those second-order desires, then, according to Frankfurt, the person is free.

[13] Frankfurt, in Watson, (1982), pp81 et seq.

Frankfurt also distinguishes between second-order desires and second-order volitions. He gives an example to illustrate this; the case of a doctor—who wants to know what it's like to crave drugs, but does not want that desire to be effective; he does not want to be led to take them. He wants to want to take drugs, but he has no desire to take them. He just wants to understand what the craving is like, but does not actually want to take them. A second-order desire is a desire to have a first-order desire, but not for it to be necessarily effective. A second order volition, by contrast, is "when he [the agent] wants a certain desire to be his will" (pp85-6)—ie., when an agent has a desire to have a certain first-order desire be his will. First or lower-order desires are the effective desires or the will. A second order desire is a desire as to how we want our lower-order desires to be. A second-order *volition* is a higher-order desire that the lower-order desire be effective.

We are free, reckons Frankfurt, if our effective (first-order) desires are in line with our second-order desires. For example, if I have a second-order desire to be generous, it means that I want it to be the case that my desire states are motivated by generosity. Thus, I would be free if I did something which was generous. It is the lower-order desires which are efficacious; not the higher. Only if the lower-order desires are in line with the higher-order, is the agent acting freely.

Frankfurt (1969, p839) draws the typical compatibilist conclusion: that someone can be morally responsible even if he was deterministically caused to do something, simply because *he may be doing what he wants*. If the will or actions go where the wants or desires go, for Frankfurt, the agent is always free. Freedom of the will is thus freedom with respect to the agent's desires (1982, p90). "It is in securing the conformity of his will to his second-order volitions, that a person exercises freedom of the will". A person identifies himself with his second-order desires (*ibid.*, p88). "He makes one of them more truly his own...". On p91 (*ibid.*), Frankfurt says:

> "...When a person identifies himself decisively with one of his first order desires, this commitment 'resounds' throughout the ... array of higher-orders."

It is because agents do this, that they are responsible for their actions.

Frankfurt uses the example of a drug addict to illustrate his model. Consider the case of a willing addict. A willing addict is a drug addict whose lower-order desires are to take drugs. We say he is willing because his second order desires—his desires about his desires—also indicate that he should take drugs. Thus, because his lower-order desires are to take drugs, and because his higher-order desires are also to take drugs, the willing addict is free—because his effective desire (the will) is constituted by, or accords with, his higher-order states. Now consider the case of the unwilling addict. His lower-order desires are to take drugs, but his higher-order desires are to refrain from taking drugs. And since the lower-order desires are efficacious, the unwilling addict will take the drugs. But because the higher-order desires conflict with the lower-order desires, the unwilling addict is unfree. An unwilling drug addict has a first-order desire to take

drugs and a second-order desire not to, and thus since he is unable to "secure the conformity" of his lower-order desires to his higher-order desires, the unwilling drug addict is not free.

Problems

There are some issues with Frankfurt's model. Firstly, it may not adequately explain the case of a willing addict. For consider: regardless of whether or not a drug addict is willing, he will be led to take the drugs *because of his addiction* rather than because of his *will*. Therefore even a willing addict would not be free, because it is not his will which has led to his action so much as his addictive desires. Even if his first- and second-order desires coincide, a willing addict is not free because it is his *addiction* which leads him to do what he does. A drug addict will take drugs *come what may* (Leon, 2001). Regardless of whether people have second-order desires to do what they do (or not), we don't think of people as being free if they can never change their minds.

Secondly, there is a concern around the causal efficacy of second-order desires. If lower-order desires are efficacious, and if our actions are free when our actions conform with our second-order desires, and our second-order desires do not *determine* the lower-order desires, then what is the point of having second-order desires? On Frankfurt's account, it seems as if the second-order desires play no significant role. Frankfurt does not say that our second-order desires control our first-order desires, so it is not clear what our second-order desires *do*. Furthermore, if on Frankfurt's account we are free when our lower-order desires conform with our second-order desires, there is no explanation for *why* there should be any such conformity, and why we would be free only if there was such conformity. There is no explanation as to whether second-order desires influence the lower-order. If the second-order desires are not able to cause the lower-order desires to be a certain way, then the second-order desires have no causal role in securing our freedom. Our first-order desires would just coincidentally conform with our second-order desires, so it would be coincidental if we were free rather than wanton.

Third, there is a question of ownership of the second-order desires. If my second-order desires were acquired through, say, a brainwashing system (or if they were just wanton—Watson, p107), it would not be clear that they would really be *my* desires, and that *I* would still be free. We need clarity on whether we as persons have any role in deciding what our second-order desires *are*. Frankfurt seems to think that it is important whether we care about what our second-order desires are. But suppose we were brainwashed to have certain second-order desires, and our first-order desires conformed to those second-order desires. On Frankfurt's model, we would still be free, because conformity of the first-order desires to the second-order is all that is required. Yet clearly someone who is brainwashed is not free. Frankfurt (p91) tries to deal with the problem of ownership of the second order by saying that we "identify" ourselves "decisively" with a "first order desire", and that this commitment "resounds through the orders". Watson describes this solution as "lame" and "arbitrary" [*sic*]. Saying that there is a

"commitment" to a certain first-order desire which "resounds throughout the orders", says Watson, is not "going to do the work Frankfurt wants it to do"—namely make the second-order desires *ours*. Watson asks: Where do the second-order desires come from? Even if we identify with the second-order desires, they could come from a brainwashing system. We need to know what the nature of the relationship between a person and their second-order desires is, in order to decide which desires belong to the person (Watson, p108). To put the question another way: What makes a second-order desire *my* desire? How is it "mine"?[14] Watson says (p149, *ibid.*) the problem is that higher-order desires and volitions lack any particular justification as to why they are not wanton, or of dubious origins; it's not clear that they belong to me.[15] There is an answer to this, but Frankfurt doesn't provide it.

Watson[16]

Watson tries, therefore, to produce a model which improves on the work of Frankfurt. He argues as follows. We take someone to be free if his action follows his will (Watson, p96). The will, in turn, should be constituted by, or accord with, what it is that the agent *values*.

> "In the case of actions that are unfree, the agent is unable to get what he most wants, or values, and this inability is due to his own 'motivational system'." "*The motivational system* of an agent is that set of considerations which will move him to action" (Watson, p106).

To be *free*, for Watson, is thus to be *able* to do what one values, and to be *unfree* is to be *unable* to do what one values (p105).

> "The free agent has the capacity to translate his values into action" (Watson, p106). "Values provide *reasons* for action" (Watson, p99). "*The valuation system* of an agent is that set of considerations which, when combined with his factual beliefs ... yields judgements of the form: the thing for me to do ... all things considered, is *a*. To ascribe free agency to a being presupposes ... [he] makes judgements of this sort" (p105).

Thus, like Frankfurt, Watson's model provides us with a mechanism for identifying different sorts of states which could determine action. If the writ of the values run—ie., if the agent acts in accordance with her values, then the agent is free. If the agent does not act in accordance with her values, then she is not free. If there is a conflict between desires (lower-order or "first-order"), and the values that the agent holds, and yet the agent acts in accordance with her values, then she is free.

[14] Leon has an answer to this question which we see in the section on "tracking" accounts.

[15] Stump disputes this. See Stump, 1993.

[16] in Watson, 1982, also in Watson, 1975.

If the higher-order desires are just desires about what desires we want to have, as we have seen, then we have the concern that those higher-order desires could also just be "wanton". Watson asks of Frankfurt: *Surely one could be a wanton with respect to one's second-order desires?* (p107) "Why does one necessarily care about one's higher-order volitions?" (p108). But if we talk in terms of what the agent *values*, we can avoid the possibility of being higher-order wantons. Consider again the example of the drug addict. If a drug addict is only motivated by his desires, he is not free—because his desires do not relate to his values. If, on the other hand, he values his being a drug addict, or wants to have desire states of a drug addict, then, according to Watson, he would be free. On p100, Watson says that his model allows us to distinguish what one values from what one desires. An (unwilling) addict desires his drugs, but does not value his being addicted to them. So he is unfree. In this way, Watson tries to avoid the problem of the nature of the higher-order desires, such as those which were problematic for Frankfurt. "What one wants most strongly [eg., drugs, need not be] what one most values" (Watson, p102). Thus Watson accounts for free action, as action which is *valued*. Contrarily, actions which are not valued, are unfree (p102). If we do something which goes against our values, we are not free. Watson says: "It is not that [agents] assign to these actions an initial value which is outweighed by other considerations. These [harmful] activities are not even represented by a positive entry, however small, on the initial 'desirability matrix'" (p101). In other words, if we have to do something that we disvalue, we are not free—*because* we disvalue it.

Watson, however, recognises a potential problem for his account. He calls this the "perverse" case (1987, p150). The perverse case is the where a person acts against what she really most values. Watson, therefore, acknowledges the possibility that we could violate our values. But he does not seem to provide an explanation in his own terms as to how this would be possible (free, yet violating what one values). For recall: for Watson, a person is free when she acts in accordance with what she values. Since the perverse case seems to act freely, and freely violates her own values, on Watson's account, a perverse case would not be free. Watson, whilst acknowledging this, does not seem to solve this problem for his account.

Further problems

Let's start with the the case of a willing addict. The addict may value taking drugs, and may, therefore, on Watson's account, be free, since on that account, a person is free if he does what he values. But a willing drug addict does *not* take the drugs because of his *values*; it is not his values that are effective. Rather, the addict takes the drugs because of his *addiction*. Therefore, since it is not his values which are operative in his taking drugs, the willing addict is not free. Since it is clear that a drug addict is not free, we may argue that a willing drug addict is an example of someone who is not free, and yet who does something he values (Leon, 2002b). This represents a problem for Watson's account.

It seems, furthermore, as if we can also freely cause harm. But Watson maintains that we cannot value doing harm (pp101-2), but it is an objective fact that we can, and often do, cause harm. And Watson argues that actions which are not valued are not free. But since Watson has argued that we cannot value doing harm, (pp101-2), inflicting harm is always unfree. Therefore, we would not be morally responsible for causing harm. I regard this as problematic for his account.

A further concern is as follows. In "Free Action and Free-will", p148, Watson mentions brainwashing as a potential problem for Frankfurt. This may be a source of second-order desires. We do not think someone would be free if she were acting on second-order *brainwashed* desires. But what if agents' *values* were *also* brainwashed or introduced by suspicious means? If an agent acted on such a valuational system, we could not regard her as free, even if her actions subsequently were in line with her valuational systems. Something would need to be added to our model of free-will, to ensure that the values we have or second-order desires that we have, are produced in the right kind of way. This is the primary problem with both models thus far.

Kane reminds us of the example of the people in B. F. Skinner's book, *Walden Two* (Kane, 1996, p65) and compares it to the kind of freedom offered by these compatibilists. In *Walden Two,* the citizens are behaviourally conditioned to do only things which bring them happiness and allow them to fulfil their life aims. Skinner claims that these people are free—because they can do what they want to do. And yet Kane asks of us, *do we really think the citizens of Walden Two are free?* The answer, he says, is *no*—because the citizens are not ultimately responsible for their own goals; they are brainwashed by Behavioural Engineers (1996, p66). The point is just this: there *must* be more to freedom than just getting what you want or value, or having higher-order desires coinciding with your lower-order desires. Because we don't know where those values or higher-order desires or values came from, and if they came from Behavioural Engineers, clearly we'd not be free.[17]

There are some writers who try to deal with the problems we have seen thus far by stipulating clauses that require that we come to our desires or values in an appropriate way. But before we deal with these more sophisticated models, let us summarise the criticisms of the traditional compatibilists which we have encountered so far.

Section 2.b.2—The Nature of Free Choice—Summary

Primarily, the question for all these models, which none of them seem to successfully answer, is this: If your higher-order desires, or valuational systems, or as we shall see, your "reasons", are what demarcate *your* actions as free, how do those higher-order desires, valuational systems, or reasons, come to be exempt from the same problems that plague the lower ones? Compatibilists

[17] Kane's example from Skinner only targets some forms of compatibilism. If compatibilist freedom meant doing what one wants because of reasons, then these people in *Walden Two* would not be free.

try to characterise *what it is for an agent to be in his action*. An agent is *in the action* if the action accords with those higher-order desires, values, reasons, that the agent has. But even if we can say that an agent's actions accord with his higher-order desires, values, or reasons, it doesn't succeed in avoiding questions as to the legitimacy of the origins of those states. There is no guarantee that those states are the way an agent would necessarily care for them to be. An agent, on these models thus far, could still be free even if his higher-order desires, values, or reasons, were not caused in the right kind of way.

Secondly, I believe that there needs to be a *robust* link between an agent's privileged mental states (ie., her values, higher-order desires, reasons) and her lower-order desires. For if we want to have someone be *represented by* her higher-order desires or values, then those states must *cause* the actions—not just coincidentally *accord* with them. Otherwise there would be no reason to have these states; they would not have causal powers. Unless there is a robust causal link (or a constitutive link), it could be coincidental that our values or higher-order desires agree with our lower-order desires. For us to be free, our values or higher-order desires must *cause* our actions.

Let us now consider the "tracking" models of freedom, which deal with these issues.

Section 3—"Reasons" and "Tracking" Models of Freedom

Leon and Wolf hold similar views—in that they also describe the mechanism for freedom in terms of a specific kind of privileged mental process—but their models are not hierarchical. Their models refer to the acquisition of reasons and beliefs.[18] People, for Wolf, are free if their actions are determined by the reasons that there are. People's beliefs about what is valuable or reasonable to do must "track" what is actually valuable or reasonable to do. If an agent does what is reasonable, or what she actually values, and she is doing what she wants to do, and her beliefs and desires are produced in the right kind of way, she is free. If an agent, however, comes to have false beliefs or values, perhaps due to interference with her cognitive abilities, she would not be able to act freely. Compare this to Leon's model:

> "What needs to be captured is the idea that the free agent is one who acts for *his* best reasons, *because* those are his best reasons; the free agent is one who acts from, and because of, his unalienated or real will. ... [alienation being] conditions which *contravene* or *infringe* an agent's autonomy; ... [where] ... there is a breakdown[19] in, or departure from, normal functioning." (Leon, 1999).

An agent acquires his beliefs, desires and values from the environment. For Leon, then, an agent is autonomous if his beliefs track the evidence in the environment, and his desires track his real

[18] Velleman has a view which is very similar except it does not rely on a tracking mechanism.

[19] Examples of 'breakdowns' would be occurrences like brainwashing, drug addiction, etc. See Leon, 2002a, p422.

interests and needs, and his values track the true objective values, the "reasons that there are", as Wolf puts it. An agent is free, on Leon's model, if he comes to have the right beliefs about what he ought to do, through a normal process of acquiring the right beliefs, desires, and values. If his beliefs indicate that all the reasons are in favour of doing something, and the agent exercises his reason to do that thing, he is acting freely, because he is acting on his best reasons. If the operative desire[20] is not appropriately derived—eg., due to some breakdown in the epistemic or cognitive process—then the agent would not have beliefs which track the truth. The same applies to the agent's desires and values—they must also have their apt determinants. If an agent's beliefs, desires and values do not track their apt determinants, then, under such circumstances, the agent would most likely do something which is against his best reasons, or against his best interests, or against the real "reasons that there are". If an agent acts against his best reasons, his operative desire would not be appropriately derived, and thus, he would be unfree (Leon, 2002a, p427, 429).[21] Thus, an agent is said to *act* freely if his acts accord with his best reasons (or his real interests, or his real values). An agent is said to *will* freely[22] if his beliefs track the objective evidence that there is, and his desires track his real needs, and his values track the true values that there are. An action is free if it tracks an autonomous will. An autonomous will is one which tracks the truth, the real values, and whatever the agent's best interests really are.

For Leon and Wolf, the fact that there are deterministic processes that lead us to have the beliefs we do, does not count against our freedom.

> "It is not when my action has any cause at all, but only when it has a special sort of cause that it is reckoned not to be free" (Ayer in Watson, 1987, p151).

We do not think of someone as unfree if he acts on his best reasons and comes to achieve what is in his best interest, because he acted on his best reasons. We only regard someone as *unfree* when he acts against his best interests, or against the reasons that there are. Leon in fact argues that we *require* determinism to be free—in order to have accurate beliefs about the environment. Beliefs, desires, and values, for Leon, must all have an appropriate determination or source—viz., the real truth, or evidence, or reason, our individual good and our real needs, and the real objective values. Only if our beliefs, desires and values are so derived, would we be what Leon calls autonomous believers, autonomous desirers, or autonomous valuers. If, however, something interferes with the normal process so described, then we could be unfree. One of the forms of interference with the normal processes would be an epistemic or cognitive failure. If freedom depends on us coming to have the right beliefs, it is immediately obvious that sometimes the environment is not conducive to this. In other words, there could be ways of learning which might be liberating, and ways of learning which would not be. It is through this possibility that we account for the cases where persons are not free. If a person is misinformed in some way,

[20] ie., the desire which ultimately leads the agent to action

[21] See also Wallace, 1994, "the power to control one's behaviour... by the light of ... reasons", p7

[22] or as Leon prefers to say, "Autonomously"

and he acts on that misinformation—we generally do not hold him accountable. However, if a person has all the facts available to him, and yet he does something that goes against the evidence of what is reasonable to do, we do hold him accountable.

Thus your freedom depends on *how* you acquire beliefs (Leon, 2002a). If you acquire beliefs (or desires, or needs) in the wrong kind of way, your *Will* would not be appropriately derived; it would not be "autonomous". Hence, your resulting actions would be unfree. The reverse also applies. If you acquire your beliefs in the right kind of way, your actions will be in accordance with your autonomous will. Leon's view is that we are responsible for our actions if we choose them, when that choice originates in an autonomous will. An autonomous will, in turn, is one which originated in an appropriate causal sequence; ie., in which our mental states track their proper determinants. It is not the case that agents believe *at will*, however. For that would involve us choosing what it is that we believe. Leon denies that our belief states must also be chosen; for that would lead to an infinite regress of willing what it is that we believe, yet, it is our will which is informed by our beliefs (Leon, 2002a, p422). Rather, Leon argues, as long as our reasons (beliefs and desires) are caused in the right way, we will have an autonomous will. The action taken must be under the control of an autonomous will in order to be free, but the will itself does not have to be controlled by the agent. Leon suggests that we can, however, indirectly influence how "open" we are to the evidence, and thus we can be held indirectly responsible for our beliefs, under certain conditions. If we turn a blind eye to the evidence that there is, for example, we would be indirectly responsible for our coming to harbour false beliefs.

Problems for the "Tracking" model

There are a number of things that concern me about the "tracking" model; specifically, (a) that I believe it may lead to a regress, or face the possibility that it does not provide adequate grounds for moral accountability, (b), that it does not allow for free wrong-doing, (c) that it might be possible to freely choose to act on a false belief (d) that we might not be able to distinguish the right kind of determinant from the wrong kind, and (e), that tracking values for the sake of acting on what is objectively beneficial, requires that values are objective. I will discuss these in detail.

3.a. Regress of responsibility

The basic concern about freedom of will is as follows. A free action is one which is willed by a person who has a free-will. Our actions are a product of our beliefs and desires, our reasons. If we acquire our reasons through deterministic processes in the environment, then what we believe or desire is caused by our environment. Thus, whatever we do (as a result of our beliefs and desires being a certain way), *is actually a result of how our environment is*. Thus, it may be argued, we are not free, because whatever we do is a result of how our environment is. This, of course, is the hard determinist position, and it is argued for in the subsequent chapter.

Let's put it another way. Our actions are free *only* if we can choose (*will*) our actions. I think a compatibilist will grant that. Our actions are thus caused by our beliefs and desires, ie., by our choosing our actions. Thus, since our actions are free only in virtue of their being chosen by our will, how is our *will* to be free? Saying that a will is "autonomous" if it is determined by the "right sort" of determinants, will not make it free in the same way as an action is free if is chosen by a free-will. Surely the will can only be free if we choose it, too? This leads to a regress.

I believe that in order for us to be morally responsible for what we do, we have to be responsible for our mental states; our mental states, beliefs, desires, and reasons, must in some significantly important way be up to us to choose or modify. For if we were not responsible for our mental states which are the causes of our actions, I cannot see how we could be responsible for our actions. Thus, if to be responsible for our actions, we have to be responsible for our reasons, we would in turn have to be responsible for the causes of our reasons too—ie., responsible for how the environment is, or be able to will that we have a different *will*. This is impossible, thus, I argue, we are not responsible for our reasons, and thus, neither for our actions.

Leon is aware of this threat of an infinite regress. His argument is that we are free as long as our desire states are produced in the "right kind of way" without any "breakdowns or departures from normal functioning". As long as the antecedent states are "appropriately caused" and the choosing "follows an appropriate path", then the choice was, in Leon's terms, "autonomous". Actions are chosen, but the antecedent desire states and belief states do not have to be chosen. Leon suggests that we are only "indirectly" responsible for our belief states. But talking about being "indirectly" responsible or how "open" we are to the evidence, won't do. That will not give us a robust form of accountability that we need for praise- or blame-worthiness. To be responsible for something, as far as I am concerned, means to have chosen it. To be indirectly responsible, I think, means to have allowed something to eventuate. Leon's model of indirect responsibility for how our belief states are, seems to be a model of negative responsibility. Just as we may be responsible indirectly for not helping someone who is drowning, so we may be indirectly responsible for believing false or bad propositions, also through neglect—through neglecting to search out available evidence when we could have done so. The point I wish to raise, in my dispute, is that regardless of whether the responsibility is positive, through active choices, or whether it is negative, through neglect of certain activities, the responsibility remains, and with that comes the implication of free choice. You chose to ignore the person who was drowning—for whatever reason. And you chose to ignore the evidence. But, if Leon is right, and we did *not* choose those belief states and desires which fully determine our actions, we could not be directly, robustly, responsible for our actions since we did not choose our beliefs. We have to be instrumental in the cause of our beliefs and desires; it is not enough for us to just have them. Merely being possessed of a state (eg., a belief or desire) does not make us responsible for it.

3.b. Freely doing wrong is the same as performing a mistake

It seems that Leon and Wolf are arguing that it is precisely when we have a cognitive failure, that our acts are unfree. Thus, conversely, when we act freely, it is because our beliefs have accurately tracked the truth as to what we ought to do. But then, Leon acknowledges, our free-will might be prone to "epistemic luck" (Leon, 2002a, p425): how moral we are or how rational our decisions are, may just depend on whether we are lucky enough to have our beliefs "track the truth" (Leon, 2002a, p426), or the evidence, or the real values, or our real interests. Suppose that the truth of what we ought to do, the truth of our best interests, etc., could directly be derived from the empirical world. In such a circumstance, we'd have to be rational and good, for recall: on this model, we do not choose our beliefs and values; they're derived from observing the environment, which is *given*. As Wolf says, we have no choice about believing the objective evidence that there is (Wolf, p61). If the environment is objective, and it determines that agents come to believe the truth, it would not be possible for us to be mistaken about what the right thing to do, is—unless we had a cognitive failure of some kind. *Ex hypothesi,* we'd only capable of good, rational, free action; for if the environment determines our beliefs, and truth is objective in the environment, then our beliefs would always be true. This is not to say, however, that our beliefs can't but track the truth, for we could experience cognitive failures. But then if we only had the wrong beliefs and desires because of a cognitive failure, surely we could not be held responsible for such a failure? Who would want their cognition to fail? No-one, surely. Thus, we could not freely do wrong; all wrong-doing would be due to cognitive errors, mistakes.[23] Given that we *can* make mistakes, and given that we *can* do things that are wrong, we must assume that sometimes we do not accurately come to believe the reasons that there are, in the environment. Accurately coming to believe the truth about what we ought to do leads us to do the right thing, so, doing the wrong thing must be a result of a mistake, or an environmental failure. Doing the wrong thing must represent a case of "a breakdown in, or departure from, normal functioning" (Leon, 1999). All wrongdoing would be accidental or mistaken, not something we could be blamed for.

Ultimately, the question of free-will is about securing responsibility. If someone does something that is against the best reasons, he/she is unfree, or has made a mistake, and therefore cannot be culpable.

Consider, using Leon's model, how we would try to explain culpable wrongdoing. Assume, in the first instance, we are dealing with someone who has accurately come to believe the truth about the environment, as to what the right or best thing to do, is. Would such a person be able to do wrong? It seems impossible. For she would have to act against the reasons that there really are, against *her best reasons*. If she has accurately come to believe the truth about what she ought to do, she can not have made an interpretative error about the environment: she knows what she should do. The fact that she does something other than what she should do, suggests some other kind of mistake or malfunction. That is the only way to explain her wrongdoing. But

[23] This view—that we are free when we do what is beneficial to do—is reminiscent of Plato's—"No man voluntarily pursues evil, or that which he thinks to be evil." (in the *Protagoras*).

are people to blame for mistakes or malfunctions? I don't think so, but Leon feels that they might be; it may depend on how "open" they are to "the evidence"—this is however only one aspect of what we can be indirectly responsible for. Now consider another possible instance. Suppose someone fails to accurately come to believe the truth about what she ought to do. Suppose the environment's deterministic influence were such that the agent nonetheless came to the wrong impression about what she ought to do. In other words, the environment provided incorrect information. Again, it is not clear that the responsibility for this failure lies with the agent. So, as we see, the only way to explain agents' doing wrong, is to assume that there was some kind of "departure from normal functioning". In which case, the agent is exempted from moral responsibility. And I think that these *are* the conditions in which we exempt people from blame.

Therefore, we cannot have a strong account of free wrong-doing on the tracking model. For any failure to do the right thing, would be explained by reference to epistemic or procedural breakdowns, or departures from normal functioning. We could only be responsible for doing the right thing, or the thing that was truly in our best interests. In those cases where we do the wrong thing, or that which is not in our best interests, we would not be morally responsible. Yet surely we want a model of free-will in which criminals and other wrongdoers can be held to account? It seems like the tracking model cannot provide this.

3.c. Are we really rational?

It is important to understand the difference between what the goal of this dissertation, and the tracking model, is. The tracking model, as it has been portrayed, aims to explicate what free-will would consist in, if it were real. This dissertation, however, asks the additional question of whether free-will exists at all, not merely what it would be if it did exist. Since the tracking model relies on the notion that human beings are rational, and do what is in their best interests, for their best reasons, the tracking model faces an empirical challenge, namely, whether humans *are* in fact rational. For if there was evidence that we are deeply irrational, that would mean that while the tracking model of freedom could be correct in its characterisation, it would suffer from a measure of practical unattainability to the extent to which we were irrational beings. For insofar as we were irrational, so we would not meet the conditions for freedom.[24]

Consider this psychometric evidence. On pp102-106, pp33-5, 120-124, Double gives some examples of experiments in which subjects were asked to write essays in response to views expressed. The experimenters found that subjects who were given positive reinforcement for attending to new views or changing their views, were the ones that changed their views. More

[24] In his *Treatise* (p460 et seq.), Hume presents us with a claim that our "reason" is merely a servant which finds ways to obtain certain objects of desire. Hume's argument is that everything we choose to do is caused by emotions. He further claims that since reason lacks emotional attributes, the reason cannot motivate us. If this could be well-substantiated, this argument would also indicate that we are not free, because we are not rational.

difficult subjects, who clung to their views, were found to do so because they had initially been lied to about the correctness of their views (ie., despite eventually being told that they were in fact wrong, they persisted in the initially reinforced belief that they were right). The evidence presented in these experiments suggests that people are not neutrally rational and *are* prone to suggestion and mistakes about the "evidence". Double cites further experiments (pp120-124) of cases where experimental subjects reported "feeling freer" if they liked the prize at the end of the tests, or if they were presented with more options during the test, or if they could base their choices on idiosyncratic features of their selves, or if the decision was easy. Stich (p9) gives similar examples with the same implication. In his examples, people persisted with beliefs they were given praise for, rather than beliefs that were rational. These experiments seem to indicate that people are suggestible, and that reason may be less relevant in helping people make decisions, than we may *prima facie* assume. If this evidence were true, it would imply that the argument of the tracking model is at least in part conditional: *If* our beliefs are produced in the right way, we would be free. This counter-evidence we see here, however, is suggesting that the beliefs will not always be produced in the right kind of way because we are not *always* "open to the evidence".

Stich (1990) presents two arguments which are relevant to this consideration of the extent of human rationality. The first is the question of the *extent* to which we could be irrational. The second is the question as to whether we have *evolved* to be rational. Both discussions could be used as firebreaks to halt the potential damage which could be inflicted on our free-choice, by a notion of our being radically irrational.

Firstly, let us consider the possibility that we may have a limit to how irrational we could be. If we could *not* be completely irrational, we could, presumably, at least sometimes, be free and morally accountable. Stich provides a summary of an argument to this effect (p50 et seq). We apply what he calls "the principle of intentional chauvinism", or "principle of humanity". That is: all people are reasonably similar and have similar intentions, beliefs, and reasoning abilities. We are all adequately or "reasonably reasonable". So, anything like us, eg., other people, should also be adequately or reasonably reasonable. Thus there is a limit to how irrational people could be; it is quite curtailed (p51). Stich, however, disputes this argument (pp51-4). The details of his disputation are not important for our purposes. The important point he makes is this: it all depends on how similar people are to us as reasoners. For people could in principle be endlessly different from us, better than us, or worse than us, as reasoners. So, necessarily, Stich rejects the concept of a normal rationality and any *a priori* argument to curtail our potential irrationality (p54).

Secondly, we consider the arguments of evolutionists, such as Dennett (in Stich, p55 et seq.). The argument has two threads. Firstly, if we were consistently irrational, and our beliefs did not track the truth about the environment, we would consistently run into trouble, probably die, and therefore be unable to perpetuate our species (breed). The fact that we have survived for this long indicates that our survival tactics, and our beliefs about our environment, must to a large extent

be accurate. If we misjudged the environment consistently, or harboured falsities about it consistently, we would have long since died out. Thus most of our survival and reasoning strategies must be rational or correct (p55). But, Stich counters, there's no indication that it is impossible for an irrational system to survive. Nor is there good proof that reasoning is an innate system, rather than learnt cognitive strategies (p56). To put the point another way: irrational or arational beings can survive. Our survival may be a matter of strategies which are based on instinctual programming, rather than strategies derived from *reasoning*. It is argued that "true beliefs" better enable us to cope with our environment (p58). But, in nature, false beliefs might let us survive. Consider Dennett's own "Intentional stance". We survive by *assuming* (often falsely), that all beings and entities have intentions about us. We can, therefore, better survive (p62), by assuming, falsely, for example, that danger is *always* present, even when it is not. Therefore, we could survive because of a false belief. Stich calls it "risk aversion". Truth, therefore, he says, is not what counts in evolution and reproduction. What counts is just survival, by whatever means, even falsehood. Of course this does not invite the conclusion that most of our beliefs ought to be false in order for us to survive, just that sometimes they do need to be; and therefore, we cannot say that all our beliefs must always track the truth.

The second thread of the evolutionist argument is that evolution produces optimal designs. As creatures evolve, their design gets better—more attuned to survive in the environment. For if this were not the case, we would not survive. In order to survive, we must be able to better survive each prior generation, and be better adapted to cope. Therefore, our rational skills, which enable us to make decisions, must be optimised for our survival (p63). This is correct. But that does not mean that our rational skills yield truths, as we have seen. Cautious beliefs, which enable us to survive, need not be true beliefs. This discussion implies that sometimes we may prosper if we acquire or hold wrong beliefs about the environment. Of course, this is not necessarily generalisable, but the fact that we can survive better, in some cases, with false beliefs, implies that it is at least not always the truth of a belief that matters, so much as its ability to allow us to prosper. Stich concludes that indeed we do not have to be rational to function in the world, and our beliefs do not have to track the truth. Our best reasons for acting may sometimes need to be false. But the "tracking" models of freedom require that our belief states track the truth. If our beliefs are sometimes false because they fail to accurately track the truth in the environment, it would mean, on the tracking model, that we would not always be free. Consider a case where I believe that I will be mugged if I go down a certain dark alley. And suppose I can see there is no-one in the alley. Suppose also that if I were to go down the alley, nothing would happen. I therefore have a false belief about being mugged. Suppose I choose to not go down the alley. Does this mean that I am not free because I have a false belief? Would the tracking model suggest that because I went against the evidence (there was no-one in the alley), that I did not freely choose to avoid the alley? Surely not. Surely we *can* go against evidence, freely?

Rather, I suspect, freedom is a matter of whether we *want* to believe certain things. Freedom of choice is about whether we can do we *want* to do, not what we have the *best reasons* for doing. The advantage of this model is that we can explain not only free good-doing, but also that it

allows us to be *freely* ignorant and destructive, which I think we are (if we are free at all).

3.d. Whether we can distinguish the "right kind" of determinants from the "wrong" kind

I am skeptical as to whether we can distinguish types of learning which are liberating, from those which are not. In other words, it is not clear that some determinants will rob us of free-will (eg., a nefarious neurosurgeon), and some will not (going to school and being subjected to the country's education system). If this supposition turns out to be correct, there would be no way to distinguish the "right kind of determinant" from the "wrong" kind, and therefore, no way to distinguish whether an action is free or not, on the tracking model. Pereboom disputes that there is a significant difference in the types of determinants that impinge upon us. He suggests that causal determinism of the normal sort represents no lesser a threat to our freedom than covert direct manipulation, as by a nefarious neurosurgeon (Pereboom, p6). Both types of determinism originate in factors which are beyond the control of the agent. Pereboom calls this "alien-deterministic" (ie., where we are impinged upon by external forces). On pp 112 et seq., Pereboom claims that he can provide a case of covert manipulation which satisfies compatibilist free-will, and that he can show how "normal" determinants collapse into the "abnormal" type. The cases that Pereboom gives make use of Frankfurt's thought experiments, but that's not important for our argument here. The question is whether Pereboom can succeed in showing that *any* determinant will rob us of free-will. In case 1, he gives an example of an agent who was created by neurosurgeons and whose first and second order desires and reasoning is produced by nefarious neurosurgeons. This agent's second-order desires are met, but he is clearly just a puppet. In his case 2, Pereboom gives a scenario in which the agent is not directly controlled by neurosurgeons, but his character was programmed by the surgeons and they created him. Thus in case 2, the agent chooses freely because his second-order desires are fulfilled, and he is getting his knowledge in the normal way, but clearly he is a puppet still, because his character was created by alien factors. In case 3, the agent was not programmed by neurosurgeons, but was created by them, and trained from a very early age to be the way he is, by his family and community, and this training happened too early in his life for him to have chosen to not allow himself to be trained. I think that intuitively, we can see this person is similar to case 2. In case 4 the situation is the same, but the agent was a normal person. Pereboom's question is: are these cases actually different? His answer is *No*. In all cases, the agent is determined by factors beyond his control, even in case 4.

So what then is the difference between the neurosurgeons and the parents, say? If we refer to the tracking model, which says that agents are free only if they come to their beliefs in "the right kind of way" we might be able to explain what the difference is between these cases. But what would we mean by "the right kind of way"? All cases provided, cases 1 through to case 4, are situations in which the agent is subject to determination by an alien source. I am arguing, in agreement with Pereboom, that there's no easy way to differentiate "normal" from "abnormal" in each case, without just being *ad hoc* about it (p116). Leon argues that the tracking model aims

to describe the conditions under which an agent would have an autonomous will, ie., conditions under which an agent would have freedom of will. That doesn't satisfy me. Even if the tracking model is correct, and we are autonomous when our reasons to act are formed "in the right kind of way", I am skeptical that we could ever know that they were so formed, or that there even is a "right kind of way". Certainly, we could be free without ever *knowing* it. But even if the tracking model preserves some sense of what it would *mean* to have freedom of will, it cannot show that we do in fact *have* freedom of will. And that, for me, is important. We should not sentence someone in a courtroom if we cannot prove they had freedom of will at the time of their misdemeanour.

3.e. Tracking values

If there were no objectively true values in the environment which we could accurately track, we could never act on true values. We could therefore not be free, in practice, unless there were such a thing as objectively true values. Since I believe there are no objective values in the environment *per se*, I conclude that we are not free in virtue of having our values instilled in "the right kind of way". There are many "values" which I would say are instilled in the right kind of way (eg., through school teachers), yet I believe that having had these values instilled in me has *reduced* my freedom, and made me more a servant of cultural determinants.

Thus, since no model of compatibilism seems to satisfy my intuitions about freedom of will, I feel we need to turn to incompatibilism, which is the topic of the next chapter.

CHAPTER 3
Incompatibilism

> Incompatibilism is the view that determinism would make freedom of will impossible (Ginet, p207). There are two forms of incompatibilism: *Libertarianism* and *Hard Determinism*. Libertarianism denies determinism and affirms free-will as a consequence, whereas hard determinism affirms determinism and denies free-will as a consequence. Both views deny that determinism could be true and that we could be free.

Section 1—The Defence of Incompatibilism

We begin our discussion with a defence of incompatibilism, which is the primary tenet of both libertarianism and hard determinism. There are two arguments for incompatibilism that will be presented here; the consequence argument and the argument for "ultimate responsibility".

1.1. The "consequence argument"

This is most succinctly expressed by Van Inwagen. He puts the argument, initially, as follows.

> "If determinism is true, then our acts are the consequences of the laws of nature and events in the past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our present acts) are not up to us." (Van Inwagen, *Preface*, i, 1983).

Van Inwagen attempts to provide a rigorous proof of this argument in his book (1983), as well as his article (1989, pp404-5). For the sake of rendering it easier to follow, I have simplified his argument slightly.

In its simplest form, the argument is as follows. (a) No-one has ever had a choice about the laws of nature. (b) No-one has ever had a choice about the distant, pre-human past. (c) The laws of nature and the distant, pre-human past determine the present (if determinism is true). (d). No-one has ever had a choice about whether the laws of nature and the pre-human past determine the present. (d) Therefore no-one has ever had a choice about how the present *is*.

To put the argument in a form closer to the original as it appears in Van Inwagen: L are laws of nature. P is a proposition which denotes some present circumstances, and P_0 is a proposition in pre-human times denoting the sum of all states of affairs. N means "it is necessary that, or no-one has ever had a choice whether". p and q are other arbitrary propositions which we use for setting up an analogous argument to establish the form of the final argument. The argument is, abbreviated and symbolised:

$N(P_0 \ \& \ L \supset P)$,
$N(P_0 \supset (L \supset P))$,
and given $[\ Np \ \& \ N(p \supset q)\ ] \supset Nq$,
and given NP_0, NL, $N(L \supset P)$,
NP

This can be read: No one has ever had a choice about (a) the circumstances prior to persons (P_0) nor (b) about the laws of nature (L), nor that (c) L and P_0 entail the present circumstances (P). No-one has had a choice about that distant past (P_0), since no person existed in those times. And no-one has had a choice that in the distant past, the laws of nature entail P. In other words, all circumstances P are only explained by the laws of nature (L) and the distant past (P_0). Now suppose *p* entails *q*. If no-one has a choice about *p*, and *p* entails *q*—and no one has a choice about that entailment either, then, it follows, no-one has a choice about *q*. Now, following that same argument form using the same logical structure, recall that no-one has ever had a choice about the circumstances in pre-human past (P_0). And no-one ever had a choice about the laws of nature (L) since they also pre-date humanity. And suppose that no-one has ever had a choice about the fact that the laws of nature determine the current circumstances P (ie, that L entails P). Given that, it is clear, that no-one has ever had a choice about P, where P is any current circumstance. This argument targets the concept of alternative possibilities. We discussed the concept of alternative possibilities in more detail in the previous chapter, so this discussion below will be brief.

A compatibilist may respond to this by saying that one could still have choice even if things are antecedently determined by the laws of nature and the distant past; for, the events could "go through the will of the agent". An agent could still have choices and alternative possibilities, for if she *had* willed otherwise, she would have *chosen* otherwise. But Van Inwagen is arguing that *if* determinism is the case, then the laws of nature and the distant past are jointly sufficient for the event (eg., the choice-event). The choice is still the agent's, but it was not up to her to do otherwise. The event will occur—come what may. The agent's will *itself* is caused by antecedent states of affairs and the laws of nature. The agent's will *itself* will always be a certain way, come what may, because of those antecedent factors. It is irrelevant whether the agent is complicit or antipathic to the event. It is irrelevant that the agent's will is subsequently causally efficacious. The point is just that if determinism is true, then the intention of the agent's will is *given*, its being complicit or antipathic is *given;* it is laid down antecedently by the laws of nature and the distant past. *What* the agent wants is not *up to* the agent; the agent's choice itself is up to L and P_0 alone. The agent does not *own* her will.

1.2. The Requirement of Ultimate Responsibility

Kane says there are two important aspects of free-will—viz., *alternative possibilities* (AP) and *ultimate responsibility* (UR). For incompatibilists, freedom means not only AP (alternative

possibilities) (Ginet, 207), it also means UR (Kane, 1996, p75). Kane argues that too much focus has been placed on AP, and too little on UR—and it is primarily the UR sense of free-will that is incompatible with determinism (2002, p224). There is more to free-will, says Kane, than just being able to do what we want or refrain from so doing at will (1996, p38, p60)—we must also be ultimately responsible for, or be the ultimate creators of, our own goals and *selves*. We must *own* our choices; they must be ours alone. We must be able to choose *what it is* that we want to do. This argument relies on the intuition that *if* it is not up to us to choose what it is that we want to do (and in so doing, build who we are as people), *then* what we *do* is not up to us either. In order to be responsible for something, we must have been *ultimately* responsible for it; the causal chain must rest at us and not be explicable by anything prior (Kane, 1996, 2002). For if we have explanatory recourse to prior events, incompatibilists feel that this would undermine our responsibility. To be responsible for some event, an incompatibilist believes we must be the sole cause of that event. Responsibility belongs to just that person who originated the event in question. For consider: We are held responsible for those things that we choose. If we did not choose to do something, but were rather "forced", we are generally not held accountable.[25] Thus, the incompatibilist believes that we are only to be held responsible for our *own* choices. The problem comes in with the *origin* of our choices. If determinism were true, then our choices would have antecedent causes which explained why they are the way they are; we would not *own* our choices. If determinism is true, then my mental states, desires, and choices, are not caused by *me*, but rather by some other factors, pre-dating me, over which I have no control. In this circumstance, an incompatibilist believes I could not be held responsible for my choices, since they would have not been *my* choices to make, rather, those choices would have been set up by the antecedent causal factors. Thus, for an incompatibilist, in order to be responsible for our actions (and choices), it must be we alone, as agents, who choose them, and they must have no prior antecedent causal factors to explain them other than *that* they are *our* choices. *To be responsible for something is to be its cause*. If we did not cause our own desire states, we could not be responsible for them; and yet it is our desire states that lead to our actions. Thus if we are not responsible for our desire states, or how open we are to evidence, etc., we cannot be responsible for our actions. It is *this* difference in definition of "free-will" that causes the disagreement between the compatibilist and incompatibilist camps.

The questions for an incompatibilist, then, are: Firstly, *is* there pervasive causal determinism, and secondly, how are we to fix responsibility within an individual agent? The following section is a discussion of various ways incompatibilist writers have attempted to deal with these matters.

Section 2—Libertarianism

Libertarians generally argue that free-will is incompatible with determinism (because of the requirement of UR), and since determinism is not the case, we must be free. I will divide the discussion of libertarianism into four related parts:

[25] We discussed the concept of alternative possibilities in more detail in the previous chapter.

1. *That morality, being self-evidently true, entails the truth of free-will*
2. *Special kinds of indeterministic agency could give us freedom*
3. *Alternative Possibilities and freedom come from causal gaps*
4. *Self-creation is the essence of free-will (Kane's theory)*

We will now proceed with a discussion and evaluation of these four approaches to providing a libertarian defence of free-will.

2.1. *That morality, being self-evidently true, entails the truth of free-will*

This somewhat strange but interesting argument in defense of libertarianism, appears in Van Inwagen, p153 et seq., 1983. It can be summarised as follows.

> *We have moral responsibility of some kind. Since moral responsibility requires free-will, we* must *have free-will. Since free-will and determinism are incompatible, determinism must be false.*

Support:

1. Free-will and determinism are incompatible.

2. It is apparent that we are at least sometimes morally responsible for what we do. It does seem that we can do things which have moral significance. And, as the compatibilists observe, it is apparent that our deliberations or thoughts and plans make a difference to what we do. We can make plans for what we want to do, and carry them out, and we are efficacious; our plans do sometimes work out. But in order for our plans to be effective, the writ of our *will* would have to sometimes run. We could not be morally responsible if our actions were not up to us. Since we are morally responsible, and since our plans are sometimes effective, what we do must at least sometimes be up to us. Therefore we are at least sometimes free, since (a) we cannot be morally responsible if we are not free to choose our actions, and (b) since our deliberations do make a difference to what we decide to do.

3. It is at least sometimes true that we have freedom in the sense described above at (2), but it is not necessarily true that determinism is the case. Thus we accept the more probable hypothesis that we have freedom, and that determinism is false.

This argument is interesting, but since I find it less than persuasive (as it contains too many assumptions), I will be brief in my treatment of it. To counter this argument, we could take a number of approaches. We could defend compatibilism and hence deny premise (1). We could deny the objectivity of moral truths and hence, moral responsibility (Double, p151 et seq.)—this

would warrant a rejection of the supporting argument at (2). Furthermore, it is possible, as the compatibilists observe, to be causally efficacious *qua* persons under a deterministic world-view. Which means we do not need indeterminism to make (2) true. We could also just argue that the appearance of free choice or moral responsibility is an illusion or misguided in some way. We could also assert that there is strong evidence for the truth of determinism and hence reject the latter part of (3)—or, if there were indeterminism, it could be that it is not occuring *where* the libertarians need it (in the right stage of the decision process). Furthermore, if compatibilism were true, (3) would be false, since it depends on (1) which would be false.

2.2. *Special kinds of indeterministic agency could give us freedom*

One of the most common argument styles relied on by libertarians—especially Taylor and Reid, as well as Chisholm (p52 et seq)—posits a special kind of causation—known as "agent-causation". This is an attempt to obtain UR. According to this view, agents not only have the ability to do what they will, but also that their wills are decided by nothing other than themselves. Actions would not be free if caused by some prior external or deterministic factor, thus actions must only be caused by the agent who owns them. Rational choices can cause actions, without being necessitated. It is merely a matter of exerting the *power to determine the will* (Reid in Lehrer, pp255 et seq.). Agent-causation, then, is a kind of causation in which an agent can spontaneously cause a mental act or decision, without himself (the agent) being antecedently determined to have those mental states (Kane, 1996, pp78, 118-9, 120-1).[26] Note that "indeterminism", here, really just means that *some* events are not *determined* by prior events (Kane, p232, Ginet, p218). This does not, however, mean that acts are not *caused*. They could still be caused: *by an agent*. Chisholm (p56) even goes so far as to say that our actions are not caused by our desires, but just by our *selves*. Nagel, who is himself not a libertarian, characterises the sense of libertarian freedom as follows: *By acting* I make one of the alternative possibilities actual, and the final explanation of the action rests with me and my choice. It is only *I* who am the reason for it, the whole reason, and no further explanation is needed (1986, p115).

The problem with this view, obviously, is that there's a lack of clarity as to how an agent could be said to have controlled or caused her action if she herself is immune to antecedent causal factors, or rather, is not herself caused to act. For surely the factors that surround the agent impinge on or cause her to have certain beliefs, which in turn cause her to have certain inclinations to act a certain way? Kane is at pains in his work to divest himself of agent-causation in this classical sense (2002, p227), and he suggests an alternative approach which we will discuss in detail later on. Let us briefly mention some difficulties with agent-causation. Firstly, it seems to take the agent out of the normal causal world. In denying that there is *any* cause for what an agent chooses other than *that* the agent chooses it, we would have no explanation for what an agent chooses, or why she chose *just that*. If the agent does not choose for the reasons that she has or the reasons that there are, there is *no* reason for the agent's

[26] Recall our inclusion of the term "spontaneous" in the commonplace Oxford definition in Chapter 1.

choice. The agent is out of the world of cause and effect—for no cause can lead the agent to choose one thing over another. Without causation, there can be no explanation for why anything happens at all. Furthermore, how does agent-causation put the agent into her actions? Is agent-causation perhaps suggesting that the only place we allow for causation is between agent and action? How is this not contradicted by science? And how is the choice that an agent makes not merely arbitrary? For surely an action is only chosen for reasons, ie., not arbitrary, if it is determined by the reasons that there are? Yet libertarians seem to want to not allow for antecedent determination of the agent—or her reasons—by antecedent causal factors, such as reasons, evidence, truth, etc. If an agent is not acting for reasons in this way, an agent is acting capriciously. This hardly seems to give an agent any ability to do specifically what is in the agent's best interests, since the agent's real best interests cannot determine the agent's choices. In other words, if the agent is herself choosing her own reasons, without there being any objective independent reasons for those choices, then her reasons (beliefs, desires, values, etc.) are not necessarily going to be those which are in fact in her best interests. They'd merely be random or arbitrary.

Let us consider O'Connor's account (p198 et seq.) which is an attempt to improve on the traditional model of agent-causation.[27] O'Connor's view is not that agents can *be* causes (as that may involve determinism), rather, his view is that agents have a kind of "probabilistic" (O'Connor, p197) ability to *steer* the choosing of an action. The choices are there, they are "given"; but we have freedom to steer one way or another. We do not originate or create actions *ex nihilo*; rather, they are things which we have reasons for, and those reasons are given in the environment. O'Connor holds that freedom is not merely a matter of how choices originate indeterministically; (*ibid.*, p201)—though our choosing *is* probabilistic—it is more than that: because agents have choices between options that they *have reasons for*. Reasons are *given*, and freedom is just the choosing of which reasons we wish to follow. We are free to choose between reasons that we have, to steer our choice one way or another between the choices we have reason for, but we are not free to choose what reasons that there *are*. The probabilistic aspect is not O'Connor's major point, however; as he reminds us, we do need *control* over decisions, not merely open possibilities (p197). We do not introduce events *ex nihilo;* we influence the direction of the choosing process. If things were merely indeterministic, O'Connor realises, this might be too "chancy" to found responsibility (p198). To ground responsibility, he argues, we need to be able to control our actions—where control is a causal concept. We shouldn't think of our selves as "arenas" in which factors work together to bring something about. Rather, O'Connor maintains, we are an "end-of-line initiator" of the resulting action (p199): we influence what is already going on inside us, and steer the process of choice by means of having reasons (p201). Just as a car (without a driver) would lack the causal ability to have an accident, (p200), agents are needed to steer their effects in the world (p201). The basic role of an agent, O'Connor says, is to generate an intention in line with the reasons that there are (O'Connor, p202). We are not (*contra* Chisholm, p56) completely God-like—we are not "prime movers

[27] O'Connor's account in part is a response to Kane, however, because I believe Kane's account is clearer, we discuss O'Connor first.

unmoved".[28] We do not need to think of it in such extreme terms. We must always, O'Connor maintains, refer to the reasons that "move" us to act as we do. But reasons are not *external* forces that act *on* us, rather, they are "*our* reasons, our own *internal* tendencies" (O'Connor, p203, my italics). So, although we have some undetermined processes going on inside our brains, when the moment of choice comes, we nonetheless do have an ability to steer one way or another depending on our *reasons*.

> "I suggest that we think of the agent's immediate effect as an action-triggering state of intention (which endures throughout the action and guides its completion)... But another aspect of that intention ... is that an action of a specific sort be performed *for certain reasons* the agent had at the time" (p201).

O'Connor rejects the idea that we need an explanation for the origins of the agent's choices. Rather, he says, free-will is a form of direct control, "*par excellence*". To demand that it is antecedently controlled is a demand that we explain how an agent controls his controlling, *ad infinitum*. This, he says, is absurd. Thus, our free-will is a kind of guiding control over whatever we have reason to do. O'Connor is still a libertarian, however, because he ultimately believes that the choices that occur to us are probabilistic in origin.

O'Connor *denies* that agents are influenced to make their decisions (*see above,* p203). But consider again O'Connor's view that agents just form the intention to act, in the light of reasons. Does this not mean that agents are causally influenced or determined by reasons? Agents *need* to be able to have their actions determined by reasons, otherwise we could not say that agents were acting *for* reasons. We have to now consider some other criticisms of O'Connor's model.

How is an agent to decide between two choices or reasons if she is unable to be causally influenced by external factors, in her decision-making process? O'Connor's view makes it hard to see *why* an agent would choose any particular reason over any other. This seems to make his view intransitive. But that I mean, we can cause, but not be caused, we can influence, but not be influenced. The causal sequence only goes out from us. It is unidirectional. But how *could* agents be able to cause effects in the world, and not be influenced by causal effects imposed upon them? Do *perceptions*, and *"the reasons that there are"* not *causally* impinge on agents? Agents need to be causally influenced by the reasons that there are. This does not mean that agents should be determined to have only one choice by the reasons and evidence, but just that agents *are* causally influenced. Otherwise they would not be able to have any reasons at all (cf. Kane, 2002, p204). How could an agent "decide" what to do if she was not influenced by reasons? Clarke argues that we have to admit that reasons would be determinants if we wish to explain at all how they could move us to actions[29] (in Kane, 2002, p205). If reasons only sometimes made an impression on us, those times that they *did* make an impression are the times where we would be causally determined by those reasons, and therefore, on O'Connor's account, be *un*free.

[28] Aquinas thought God was the first cause—*prima causa* (Chisholm, p49).

[29] Clarke (in Kane, 2002, p205 et seq.) argues apparently for probabilistic causation by reasons

Ginet offers a similar view to O'Connor, but a more radical one, which is also an attempt to improve on the traditional model of agent-causation. He rejects causation of *any* kind in relation to actions, possibly because he senses that any form of causation, if allowed, would admit determinism—ie., unlike Kane, Ginet seems to believe that causation entails determinism.[30] Ginet offers a view that he calls "simple indeterminism".[31] He defines it as follows: If an agent does something that is followed by some event *e*, and *e* is a free action of that agent, then "she makes it the case that *e* occurs, not by causing it, but by simply *performing* it" (Ginet, p208). Ginet denies that this involves causation at all. Actions only count *as* actions (and are free), says Ginet, if they have an "actish" phenomenal quality, ie., it is *as if* I cause them, or it is *as if* I make some event occur. My act must have a "quality of its seeming as if I directly bring it about" if it is to be free (p210).

> "I make my own free, simple mental acts occur, not by causing them, but simply by being their subject, by their being my acts. They are *ipso facto* determined or controlled by me, provided they are free, that is, not determined by something else." (p210).

Ginet has a libertarian argument for reasons-explanations for action (as do Kane and O'Connor). Ginet's idea is that if we act for the reasons we have, and our acts are not antecedently determined, we are free. Reasons, he feels, do not entail causality. In this respect, he rejects the compatibilist criticism of libertarianism, viz., that acting for reasons entails causality.

> "If [one] were still to insist that explanation [for choice] always requires causation, then he would seem to me to be just clinging to a dogma, blind to the possibility that in reasons explanation we have a fundamentally different kind of explanation, a non-causal kind." (p218).

Unfortunately for Ginet, Science relies on causal explanations. Our current models of mental states, mental content, and what happens in our performing actions, are explained in terms of causal relationships between events in physiological matter—electrochemical events in our brains. Our "choices" involve neurochemistry, which is to be described causally. The molecules in our neurons do not move around because they themselves have "reasons" to do so (or goals or aims). They move around due to deterministic causes.

Ginet tries to dismissively appeal to intuition as if it were a *brute "fact"* to defend his view—of how it is that we can do things without causation—"Well, it just is [so]" (*sic*, Ginet, p213). Nagel (1986, p111), however, says that just being a subject of an experiencing of a volition is not enough to make it an act. "Seeming as if I directly bring it about" (Ginet, p210) is not the same thing, at all, as *actually* bringing it about. Acts cannot be *"ipso facto* controlled by me" if they are not *caused* by me. Appeals to subjective qualities of free actions as proof that they are free, is

[30] Kane, as we saw, feels that reasons and *quantum* causation do not entail determinism.

[31] Ginet denies that his view is agent-causal. Whether it is a form of agent causation or not, is immaterial, says Ginet, as long as we accept the model (Ginet, pp208-9, p218).

a weak tactic. What, we must ask Ginet, would make my acts, my choices, in any sense "mine", if they were not caused by my reasons? Furthermore, if we were relying on choices being free because they have an "actish quality" to make them mine, another problem is simply that, at most, this admits that choice has a subjective aspect. If the "actish quality", the subjective feel of a choice, is the defining property, the most important property, then choices might become epiphenomenal—a mere "feeling of being like a choice" rather than being an actual, causally-efficacious choice. While this may be what choices really are (as we shall later see), it does not do the work that Ginet wants; namely allow choices to be efficacious in the absence of determinism. Agents just "perform" actions, where "perform" means to do something yet without there being an antecedent causal event that necessitated the doing of that thing.

To summarise, the problem for all the models of agent-causal libertarianism, is that we cannot explain what counts as a choice of an agent if the agent does not *determine* the choice. If the agent *does* determine the choice at all, *then some form of determinism is true*. Libertarians of this sort argue that agent-causation is a causal model, allowing determination of choice by the agent, but not allowing for determinism of the agent by antecedent causal factors. I cannot see, however, that we can grant this to a libertarian. Either libertarians must allow that there is determinism, or they must not. If they want to allow determinism from the agent's will to the action, they must be consistent and allow other causal factors as well. If, like Kane or van Inwagen (as we shall later see), libertarians want to rely on quantum randomness and thus deny determinism—for this is what quantum randomness does deny—then they must deny causal efficacy of the agent, too. Agent-causation cannot operate unless there *is* causation, and if there is causation, we have no solid reason to argue that agents cannot be causally impinged upon by antecedent causal factors. Futhermore, if quantum randomness is meant to give agents freedom of will, so that their choices are free—then it looks as if action choices are not bound by causal laws. If this were the case, then actions would merely be random. But we do not have evidence that everything is quantum-random. It is not quantum-random that as an apple falls from a tree, it falls downwards. That is determined by the force of gravity. The same applies to persons. We might grant that persons' neurochemistry is quantum-random, but then it is only persons' actions which would be random, and not bound by causal law. If, therefore, the agent only probabilistically determines his or her choice (or "just performs it"), then we can see no reason *why* the agent chose that particular act over any other. Appeals to "reasons" won't help; for if a reason is the reason *why* an agent performed some act, then the reason *determined* the agent's choice to perform that act. "Performing" an act must mean "determining" or "causing" the act. If it did not mean that, then it would be pure coincidence that an agent wished to perform some act, and that the act came to be. An agent cannot be said to have control over his actions if his will does not prevail. If an agent's will prevails, then it must be the case that the agent's will *causally determines* what action occurs. This seems to contradict the libertarians' belief that causal determinism is not the case. Recall: Libertarianism is the view that (a) incompatibilism is the case, and (b) causal determinism is not the case. Yet they still want to allow agent-causation, which contradicts (b).

2.3: ***Alternative Possibilities and freedom come from causal gaps***

Libertarians, generally, try to locate a breach in the deterministic causal chain to obtain freedom for us. This argument can be expressed briefly as follows:

> i) Incompatibilism is true
>
> ii) If there was a gap in the causal sequence, where the antecedent conditions did not fully necessitate the subsequent events, then there would be room for freedom, since choice would not be bound by strict mechanistic sequences. If there were, in other words, no *sufficient* reasons in antecedent conditions, without our intervention through choice, we would be free.
>
> iii) There are in fact such causal gaps
>
> iv) Therefore we are sometimes free

Libertarians, generally would like unconditional alternative possibilities (UCA). They would like to be able to "choose otherwise", even if the circumstances (context, environment, personal history of the agent, options available)—were all held constant (Ginet, p207).

> "Consider... the categorical sense of 'could have chosen otherwise' that Libertarians impute to free agents. According to that sense, when I freely select one alternative, I possess the categorical freedom to choose otherwise in exactly that same circumstance. Both options are equally available to me, given my condition at the moment, and indeed, given my entire psychological history." (Double, p14).

As we saw in the chapter on compatibilism, this kind of alternative possibility was rejected (recall the Doppelgänger)—as it entails the possibility of acting against reason. However, libertarians seem to require it; so they argue that UCA is possible through the lack of causal necessitation at some point in the process of volition. Let us therefore consider some possible causal gaps.

i. Delays as ways to get alternative possibilities

On p211 Double discusses what he calls "delay libertarianism". In this form of libertarianism, the *reasons* "set the stage" for a decision process, but there is a delay before action in which a free agent could change her mind about the decision, or, some neural indeterminacy could overturn the decision. The idea is that if there is some delay, it provides a gap in which the agent has the opportunity to exercise free choice in the form of a counter-decision—a *changing of mind*. This is very much like the delay choices ("vetoes") we will see later in the chapter on Libet, where agents only have a choice to change their mind about whether to do something or

not; no more than that. The idea is this: if the inexorable causal sequence of the universe slows down (eg., inside our heads) just long enough for us to think about what we want to do, we then have an opportunity to change our minds about what it is that we want to do.

But, Double asserts: "A chance to change one's mind that is contingent on delays does not provide control over the alternative choices that are not made if the delays fail to occur" (p214). In other words, this fails to obtain UCA for us (p213)—for having delays would only give us a choice like [A or *not-A*], it would not give us a choice of the form [A or *B*]—which is what we require for *full* unconditional alternative possibilities.

To put the concern another way; delays, in and of themselves, will not provide a break in the causal chain, merely a *slowing down* of the causal chain. Slowing things down is only enough to give us a limited choice. We may have to settle for [A or *not*-A] as the best we can get as a range of choices for free-will, or unconditional alternative possibilities. Slowing things down certainly does not *remove* causal determinism.

ii. Quantum Mechanics as a causal gap

If libertarians could argue that determinism was not the case, and that incompatibilism was the case, they might be able to establish their point. This is what libertarians aim for when they look to Quantum Mechanics: they seek the refutation of determinism.[32]

According to some scientists, the behaviours of fundamental particles which make up matter, are *indeterministic*.[33] Their behaviours are probabilistic. If our minds are made up of atoms, then the chemical behaviours of our minds would also be probabilistic. Since our minds are probably supervenient (in some way) on the chemical behaviours of our neurons, our minds are most likely also prone to be influenced by quantum indeterminism (Honderich, 1993, p62). Peter Van Inwagen (1983) argues for this theory, as does Robert Kane (1996, 2002). The argument, in its briefest form, is this:

> *Indeterminism is a fact, in virtue of Quantum Mechanics. Thus since determinism is false, and since determinism is incompatible with free-will, we must have free-will.*[34]

There are some problems with this. Firstly, as Leon says in his article *On the Value and Scope of Freedom*—probabilistic or random acts, of the sort that a libertarian regards as "free", would actually just be "capricious" (p4)—lacking good reason. We want actions that express or

[32] Hume famously expressed skepticism about causation which may lead one to wonder whether determinism is real. We only perceive sequences of events, he says; there's nothing obvious about the sequences which *proves* that they are *causal* (1748, §V, Part 1, also ¶71, §VIII. "Of Liberty and Necessity").

[33] Giancoli, pp752 et seq.

[34] assuming all other things are held equal, eg., we are creatures capable of choices.

conform to our values, desires, and reasons. If agents' choices are not determined by reasons, the agents would not be acting for any good reason. In other words, they would be acting capriciously. In fact, if actions were indeterministic, agents could not be the causes of their own actions. Such random bodily motions wouldn't even begin to qualify as actions. Quantum mechanics would make us victims of indeterminate physics. On p199, Double says: "I do not control that which is indeterminate since indeterminate events are not under the control of anyone or anything".

> "If the act—the firing of the shot—was not caused at all, if it was fortuitous or capricious, happening so to speak out of the blue, then, presumably, no one—and nothing—was responsible for the act." (Chisholm, R., 1964, in Double, p13).

Van Inwagen appreciates this point (Van Inwagen, p134, 1983)—as does Kane, who spends his last four chapters (1996) trying to avoid the problem. Kane gives us the example of the assassin, whose gun fires if he wants it to fire, and whose gun may fail to fire under certain other circumstances. If the gun fires, and the assassin wanted it to fire, he was responsible. If it fires randomly, this does not matter, as long as the assassin pulling the trigger is a necessary condition. He wants to fire the shot, but whether or not it is fired, is quantum random, depending on a spasm in his arm. The important point, however, is not that the shot was fired only if there was that quantum spasm; the important point is that the assassin wanted to fire the shot. In my opinion, *if* the gun fires *only* because of a quantum event, and it is against the will of the assassin (eg., if he changed his mind), *then* he would not be responsible.

A second problem for libertarian arguments that rely on quantum mechanics, is whether the quantum states in our brains do in fact get "amplified up" to the decision level. Human action *is* regular (Honderich, 1993). It certainly doesn't *seem* to be the case that our actions are random or arbitrary. Furthermore, on p212, Double argues that it cannot be that just as we prepare to make a decision, that quantum particles "know" that they must "go on a spree" (*sic*). "The question of why quantum indeterminacies should occur just when we manifest libertarian free-will strikes me as unanswerable" (Double, p221). Pereboom points out (p84), that even if everything is quantum random, there is a statistical perrcentage probability we expect of certain types of events as determined by the quantum events' probabilities. Which means that unless agent-causal choices accurately track the probability of the quantum events, agent-causal events do not track even quantum events—which means agent-causal events are not explained by quantum events. Only if we lack agent-causation, would our actions regularly tie up with the regular percentage probability of certain quantum events. But even such things are unlikely; for example (p81), a soda bottle on a table would only move one inch aside if coincidentally all the quantum events lined up at the right instant to do that, which is highly unlikely. We therefore would need an explanation as to why an agents' choices lined up with the quantum mechanical probabilities. If, for example, we have a likelihood of doing action X of 87%, we expect that actions would follow that probability. But on the other hand, if the agent really is free, the agent has 100% freedom to do otherwise, so how are we to explain the fact that the agent also does X

87% of the time? Surely if they have true free-will they'd not have to do X 87% of the time? And if so, it means that it's not the quantum events which lead to the choices. If free agents were truly free and their actions happened to match up every time with the quantum events, this would be a 'wild coincidence', says Pereboom. Thus, if the choices lined up on the statistical chances of X, then it makes no difference if the choices for X happen with or without agent intervention, since they will line up on the 87% mark either way. Having free-will, therefore, wouldn't matter—because the resultant choice X would occur on an 87% regularity—either due to quantum mechanics, or, it would just be a wild coincidence.

But let us focus on Robert Kane's theory as an example of this kind of argument; for Kane believes he has some answers.

2.4.i *Self-creation is the essence of free-will (Kane's theory)*

Kane's model (2002, pp223 et seq.) is an attempt to deal with the shortcomings of traditional agent-causal models, and relies on quantum indeterminism. Kane argues that there is at least one form of free-will that is worth wanting that is not compatible with determinism, namely, the form in which one has "the power to be the ultimate creator and sustainer of one's own ends or purposes" (2002, p223).

> "To be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient reason (condition, cause or motive) for the actions occurring. If, for example, a choice issues from, and can be sufficiently explained, by an agent's character and motives (together with background conditions, then to be *ultimately* responsible for the choice, the agent must at least in part be responsible by virtue of the choices or actions volutarily performed in the past for having the character and motives he or she now has. ... This UR condition accounts for the "ultimate" in the original definition of free-will [cited above, 2002, p223]. Now UR does not require that we could have done otherwise (AP) for every act done of our own free-wills... [It] *does* require that we could have done otherwise with respect to some acts in our past life histories by which we formed our present characters. I call these self-forming actions, or SFAs." (Kane, 2002, pp224-5).

Kane argues that it is in virtue of having the freedom to choose SFAs through moral dilemmas and some other challenging choices which we face, that we form our characters—*and it is through our choice of who we are that we have free-will*. If none of our acts were self-forming, he asserts, we could never be responsible, because it is in the choosing of SFAs that we exercise free choice—choice that is *undetermined* (2002, p225). The only way we can avoid the hard determinist consequence of never being responsible for anything, is if we are at least responsible (as *sufficient cause)* for *some* things in our lives. If we want to avoid the regress of responsibility, we need to at least be the sole originators of some of our choices—the most important ones that form *who we are*.

How, then, could we choose an SFA? On pp226 et seq., Kane explains: Since quantum indeterminacies are uncaused by prior states, and since they can occur in our brains where our decisions are made, it is possible for us to have an event occur within us which assists in making a decision that is not determined by prior states of affairs. Kane rejects the need for any special kind of agent-causation in the traditional libertarian sense (2002, p227).[35] He also says we do not need quantum randomness all the time. For if an action is to be *ours*, it must be caused by our selves. That would require a kind of deterministic causation (2002, p227). Rather, he says, it is only in selection of the SFAs that make us who we are, that we need free choice—choice that is undetermined or quantum-random. Kane gives us an example: of a businesswoman on her way to an important meeting, who hears someone being mugged in an alley. The businesswoman has a dilemma: either continue to the meeting and help her career, or shout for help and resolve a morally bad situation. Both paths have their attractions, and she has difficulty deciding. She has reason for both actions. But it is through a quantum-random event, which allows her to go one way, rather than the other, that she eventually makes the decision. She puts an effort into making one decision, and in so doing, *by deciding,* wilfully sets one of the alternatives aside. She allows one of her efforts of will to win (2002, p228). The indecisiveness that she feels at first is instantiated in her brain by indeterministic neurological noise that places obstacles in her decision process. That neurological noise is an obstacle that she overcomes (2002, p229). But just because the randomness gives uncertainty as to which decision she will take, it does not mean she does not endorse one of the courses of action. She has reason for both. The randomness will ultimately be overcome by an act of will, and she will voluntarily go one way or the other—but which way, exactly, is indeterminate.

Kane stresses that to be responsible, we must ultimately be able to choose without prior determination (Kane, 1996, p75), ie., be the ultimate source of our own actions. If we deny this criterion, then we end up with a regress to factors beyond our control, to sufficient conditions other than our motive(s) (1996, p37, 63-4, 114). If we want to stop the regress, agents must in some way be capable of generating these "self-creating acts or willings" (1996, p124-5), which lead agents to become the people they are (1996, pp75, 79, 114, 144). He feels that somehow the *choosing* action of the mind is also a quantum random or chaotic event—and hence it is truly sponaneous, *truly not* antecedently determined, and truly originating in the persona of the agent (1996, p131, 138-40). In (2002, p230) Kane says: Agents do not control or determine which choice outcome will occur before it occurs; but they do control which of them occurs *when* it occurs. Because the agent is trying to make either outcome occur (the agent has reason for both), the agent is responsible for whichever occurs. And in so doing, the agent has performed an SFA, and will thus further create, freely, who she is as a person. According to Kane, then, self-creation is the primary way to free-will (1996, p34).

Peoples' characters are built over time, by self-forming actions (SFAs). But those SFAs have to either be caused by prior mental states—which in turn were caused by prior SFAs, or by random neuronal events. But this may lead us to determinism again. So Kane postulates that

[35] We discuss these after the material on Kane

indeterminism is involved in the causal chain between prior SFAs the subsequent choices of SFA (see note 19, p245, 2002). As long as the original SFA(s) were indeterministic, Kane is satisfied that our personalities were indeterministically created, and that we are therefore free.

There are some concerns here that Kane may be prone to the same accusations that are usually levelled; that agents' choices would be arbitrary. Kane himself articulates this objection:

> "It is difficult to see how the choice outcomes ... could be in the agent's control rather than merely matters of chance." (1996, p131).

In other words, it is not clear what the quantum-random neural noise does. If choices are in the agent's control entirely, then the choices are deterministic. If the agent does not entirely control her choice, and it is influenced by the quantum-random neural noise, then the agent loses control over her choice. He tries to settle on this formulation in order to resolve the apparent difficulty:

> "Paradoxical as it may seem, in order to have ultimate control [and hence responsibility] over their destinies, possessors of free-will must relinquish ... antecedent determining control that would guarantee how things will turn out in advance... When they engage in self-formation, what they choose is **not** determined by their *already* formed characters and motives." (1996, p144).

Kane tries to deal with this danger, that we may be acting arbitrarily, as follows. In (2002), p232, he mentions that the chief criticism against libertarians, namely, that they offer random or capricious bodily movements, instead of actions chosen for reasons, is mistaken. He argues that this criticism relies on the idea that "undetermined" actions are literally "uncaused". Kane denies this. Instead, he argues, "undetermined" just refers to a different kind of causation to deterministic causation. Kane claims that

> "... indeterminism is a technical term that merely precludes *deterministic* causation, though not causation altogether. Indeterminism is consistent with nondeterministic or probabilistic causation, where the outcome is not inevitable. It is therefore a mistake ... to assume that 'undetermined' means 'uncaused'".

Kane may be right. Peoples' actions do seem "spontaneous" (see also Ginet, pp214-5), but we also know that peoples' choices are causally efficacious and done for reasons. If Kane is correct in denying that indeterminate means random, we would not be entitled to say that persons' actions are *capricious* under libertarian free-will—because they *would* actually be caused (even if they were practically indeterminable). In fact, Kane says that we only take recourse to the quantum indeterminacies or chaotic neural events in our brains to explain why one event occurred rather than the other. So, it is not that libertarian-free actions are random, rather, they are just indeterministic. But that does not mean that they do not occur for reasons.

To summarise Kane's model, then:

- There are reasons for acting
- Indeterministic chaos in brain states allow that particular choices are not sufficient to ensure what occurs
- Neurological noise in the brain states provides internal impediments—for agents to overcome
- The agent has reason for either path or either choice
- The agent overcomes the indeterministic noise in her brain and endorses one path in particular by acting on that particular choice; choosing that one particular path
- That act is an SFA, and builds the agent's character, thus setting the course (causally), for future actions.

Sartre (see below) and Glover (p446) both make similar arguments. The idea is this: *who we are as a person, determines what we do. Therefore, to be free, we need control or choice over who we are* (Kane, 2002, p231). Sartre (p26 et seq.), argues that "existence precedes essence"—ie., we exist before we have a nature; we choose who we are to become. Thus, for Sartre, the more we can self-create, the more free we are.[36] If we were deterministically created, this would pose a threat to our project to be able to render our lives meaningful by being actively involved in choosing who it is that we want to be: "Ruling out all courses of action but one, it [determinism] eliminates genuine choice" (Glover, p454).

2.4.ii. Criticisms of Kane's Model

Whether Kane's model makes us free

I am impressed by Kane's description of how decisions take place, and think he might be right: decisions (in his case, between two courses of action that we have reason for) might be quantum-indeterminate or chaotic in their *coming-to-be*. But that these quantum-random events influence my subsequent actions is something I cannot avoid. I am at the mercy of these quantum-random events. The issue is not what *kind* of causality prevails; it is a matter of whether we have any say over what prevails at all. For us to be free on an incompatibilist model, we really need to have a say over what prevails—in the formation of who we are as well as everyday choices. That we have no say over what prevails, for me, is what matters. Having reason for either course is a good start, but ultimately, *I* want to choose, not have some quantum-randomiser make the choice for

[36] Sartre's version of the argument is problematic. He argues that we have no essence and create ourselves—and yet, he predicates this on an argument that we *do* have an essence—of inherent free-will.

me. Kane seems to be taking that control away and leaving agents at the mercy of quantum randomness. My concern here is whether quantum states *cause* our decisions, or whether they just have an indeterministic effect on our decisions. If quantum randomness causes our decisions, they are not our decisions. If choices are just a result of quantum randomness, how is the agent then to be the explanation for the action? If, on the other hand, quantum randomness just has an indeterministic effect on our decisions, that does not necessarily ensure that no deterministic factors can impinge on our actions, and thereby render our actions not perfectly freely chosen by us alone.

Kane can respond. Kane uses quantum mechanics as an *enabler* of a choice between one and another option—*not* as a cause of the choice. Quantum background noise creates an environment in which no antecedent causal factors can necessitate a certain action, but this does not mean that our choices are random. Quantum states *set us free*. They *make* free choices possible—ie., they give us AP. We don't *need* to be able to control them, for our motivation alone is a sufficient condition. In other words, even without the randomness, we have motives for either choice. In the absence of quantum randomness, motivation or reasons would still exist, and the writ of our desires would still run. Self-forming actions are *sufficiently* caused by our motives. In the absence of quantum mechanics, choices would be deterministic. Kane allows that. Choice, he says, only needs to be quantum-random in the case of SFAs—the choices which make me who I am. The quantum event does not make the choice go one way or the other, *per se*, because it is only the motive which is a sufficient condition. In other words, with or without the randomness, I'd still choose, so we're not at the mercy of randomness.

But what purchase does such a randomiser *have* then, if we would choose, anyway, without it? Certainly, a randomiser can ensure that the choice is not caused by some antecedent factor (eg., my family or upbringing)—a randomiser could ensure that the choice happened only inside me. But if my freedom depends on quantum mechanics, *and* I'd do what I was going to do anyway with or without the randomiser, all that the randomiser does is take away self-control. Probabilistic causation does not remove causation, it just removes full control from the hands of the agent. Kane can't have it both ways. Either quantum randomness is why we act, or reasons are why we act. If we have reason to do something, and it is a quantum event that leads us to choose, then the choice is taken because of the quantum event, not the reasons. If we would make either choice without the quantum randomiser, then the reasons were not equal and one of them had deterministic causal power. If the reasons were equal, we'd not make the choice at all without the randomiser, in which case, our choosing any of them is random, not because of the rational pull of one over the other. Either we are pushed over the threshold of choice by reason or by the randomiser. If the randomiser is the necessary factor in our choice, the randomiser is ultimately why we chose. If the reason is the necessary factor, then reason is ultimately why we chose. What we care about is doing things that we have good reason for, not just arbitrariness that we ourselves have no say over.

Without reference to prior causes, furthermore, it is still mysterious as to why one choice is taken rather than another (Double, p207). If character, motives and reasons do not determine the choice (because determinism is denied), then what does? If quantum randomness is something to "overcome", how does that make the action "free"? What, in the context of quantum noise, would lead an agent to favour one choice rather than another? Why would a *particular* effort be taken, and how are we to explain *that* effort? (Double, p209). Real responsibility requires not just "ownership of effort", it requires "imperialistic control over choice" (Double, p210). Surely the freedom, for an incompatibilist, comes in when the outcome is *not* inevitable, when it is steered *by* a quantum spasm? And if that's the case, the quantum spasm *is* a necessary cause for my choice to be free. Let's refer back to Kane's argument again. Suppose the cause of a choice is my motivational states. Suppose that with or without the quantum states, my motivational states would follow through in action. The quantum states, as we can see in this picture, do not contribute causally to my choosing. If they make the choice go a different way, that can only mean that there were multiple motivational states *and* the quantum states contributed causally to ensure only one particular motivatory state followed through. In the absence of the quantum states, we would assume, that since in the quantum state, motivational state *A* followed through, it must have been the strongest motivational state because it overcame Kane's quantum noise. Now in the non-quantum state—ie., where there is no quantum noise—that same state *A* would *still* be the strongest, and would still follow through. I can only conclude that either the quantum states *do not* contribute causally to my choosing anything (in which case they do not provide alternative possibilities)—or, if quantum states *do* make a difference to my choices, then they *must* contribute causally to those choices. If they contribute causally to one of my choices, even if I like the choice, it was not fully up to me to make that choice, in which case I would not have had self-control. The problem with Kane's model, so far, is not one of caprice, rather, it is that he seems to be stripping us of control. If, as he admits, we cannot control the antecedent quantum events (Kane, 1996, p144), and if he admits that actions lacking antecedent quantum events are not free, then either we don't really control our SFAs or they're not free either.

Why, furthermore, is Kane assuming that in SFAs, we always have reason for both our choices, and the ultimate choice is always quantum random? I can think of plenty of choices in my past, which were defining moments for who I am, which were very quick choices to make, not involving any struggle against quantum noise, and not involving two options that were of almost equal merit. While it *is* true that for most choices we approve of either option in a choice *only partly,* not *all* choices are like that. In many choices, we don't approve of one of the choices *at all*. Decisive choices are the quickest choices, *where we feel freest* (Double, pp120-124). I bet no quantum indeterminacy is needed to push one of those choices through—but rather just plain old *predictable* determinism. If Kane wants to predicate free-will on quantum indeterminacy (which he often asserts), we would find that the quickest choices, where we feel *freest*, are the ones which are *determined by our selves*, and the ones where we feel least *under control*, *least* free, are the ones that take some rumination and (if Kane is right) random quantum interference

to be pushed through.[37] If my character *determines* a choice and it follows through in action, and I get what I want, then I will feel free—because *I* chose it. If I am faced with a dilemma and eventually a random brain event makes me choose, I will feel unfree—because it wasn't *me* who wanted to make the choice; it was a choice which arose in me in some way. Thus determined decisions at least will *feel* freer than quantum-random decisions, which would feel unfree.

Causa Sui Fallacy Threatens

Kane says we are free when our actions are SFAs, or chosen by our selves (and our selves are formed by SFA choices). SFAs build who we are—they build our *selves*, and our selves make our choices. My concern here is this. Supposing my first choice was free, and was an SFA. That would mean that my first choice (ever, as a person), was chosen by my self. But if my first choice ever was chosen by my self, it would mean that my self existed before I had ever made any SFA. Therefore, the self pre-exists SFAs. Therefore the self is not created by an SFA. The only alternative is to suppose that I could have an SFA prior to having a self—ie., that I could freely choose who my self was to become. But we have already seen that choices come from the self. So I could not choose who my self was to become, because I'd need a self in order to choose my future self.

> "The *causa sui* is the best self-contradiction hitherto imagined... For the desire for 'freedom of will' ... is nothing less than the desire to be precisely that *causa sui* and ... to pull oneself into existence out of the swamp of nothingness by one's own hair." (Nietzsche, pp50-51, *Beyond Good and Evil*).

Self-creation is a logical impossibility, and thus cannot be a project of humanity, contrary to what Sartre and other libertarians say (G. Strawson, p5). To be responsible for one's actions one would have to be responsible for the way one is mentally. That is, one's mind would be responsible for its own self-creation. If an agent's effort in choosing an action originates in his character, and his character came to be because of prior factors beyond his control, whether random or causally determined, his character is not something he created. Therefore whatever effort he makes in choosing, it is beyond his control. Even if his choices, as Kane says, partially create his character, the point is that those choices are based on prior choices. Consider the first choice an agent makes: it was not completely or ultimately up to him (Pereboom, pp48-9).

Libertarianism: Conclusion

We have rejected randomness as a source of freedom. As Kane rightly points out, we can only make someone responsible for his acts if he is ultimately the sole cause of those acts (Kane, 1996, pp37, 63-4, 114). If agents are responsible for their acts, they *causally determine* their acts. But Kane argues that our choice between one or another path of action is a matter of

[37] Please refer to Double's mention of experiments (pp120-124) for evidence that quick decisions feel freer.

chance, even if we have reason for both choices. We have argued that "probabilistic choice" actually involves dropping some self-control, and that the best a quantum-random choice mechanism will do is remove some of our personal involvement in choice. If our acts were quantum-random, we'd have no ability to direct choices, either—either they'd be random, or we'd have no explanation for *why* we chose *what* we chose. Random events leading to choices do not provide the sort of control over our decsions that moral responsibility requires (Pereboom, p53); random events just make the outcome random; they don't put control back in the hands of the agent (*ibid.*, p55).

We have also rejected self-creation as a logical impossibility. A self cannot decide what it is going to become unless that self already exists. Thus since the self already exists prior to decisions, the self must be created by things other than itself. Therefore *who we are as persons* or selves, is not up to us to choose. But our choices are caused by *who we are as persons*.

Finally, mysterious forms of causation, eg., agent-causation, have been rejected as well. Agent-causal models do not show why we act, or provide mechanisms for our actions to be connected to objective reasons to act. Agent-causal models might preserve our ability to choose and not be antecedently determined, but then we would be choosing capriciously, because we would not be led to choose by the objective evidence and reasons. Similarly, after we had chosen, unless we *caused* our own actions, we could not be responsible for our actions; because to be responsible for an event is to be its cause. Thus agent-causation prevents us from acting rationally or according to evidence and reason, and it also prevents us from being responsible for what we do or choose, and breaks the important causal tie between us and our actions.

Thus, as we can see, it looks like there are substantial problems with libertarianism. This leads us to consider the other form of incompatibilism—hard determinism—which may yield a result that copes with our intuition that we need to be "ultimately responsible" (Kane, 1996, p75).

Section 3—Hard determinism

Hard determinism is the view that (a) determinism is the case and (b) that determinism is incompatible with free-will. *Assuming* we have an adequate proof of incompatibilism, and *assuming* the reader agrees that determinism is true (for how else could we explain anything if everything were undetermined?)—we have provided the prerequisites necessary to come to the hard determinist's conclusion. No matter *how* our "decisions" originate (deterministically *or* probabilistically), the point is our decisions are not ultimately in our control (Nagel, 1986, p115, p123). Our decisions are not up to us: they occur because of our society, parents, genetics, peers, etc. Thus we can never be ultimately responsible for who we are and what we do (see also Honderich, 1973, p187). We could leave the chapter at this point, however, there are some additional considerations that favour hard determinism, that are worth mentioning.

Psychosocial Determinism

Many writers assume that the argument for hard determinism is based on neurological determinism. The argument is as follows:

> *Choices are mental events. Mental events depend on brain events. Brain events are neurochemical. Neurochemistry is deterministic and governed by scientifc laws. Therefore choice is deterministic and governed by scientific laws.*

For a more elaborate defense of this argument, see eg., Honderich, 1973, p187.

But hard determinism does not have to rely on this kind of argument.[38] We can base it on a variety of other explanations, including psychological determinants (eg., Freud's doctrines, in Sdorow, p506 et seq.), or environmental determinants (eg., Locke[39]), or a combination. We might also make reference to genetic or hereditary determinism.

The idea of psychosocial determinism is not that, as in Psychology, the discipline, or Sociology, we are trying to come up with laws or law-like formulations that describe or predict human behaviour. The idea of this model of determinist explanation is to arrive at a description of the operation of causality on our choices; and this model is agnostic to the truth or falsity of the doctrines of Central State Materialism, Supervenience Theory, or whatever other materialist theory of mind one wishes to name. Psychosocial Determinism is a theory which even Descartes, and anyone else who believed in free-floating souls or minds, could accept. In this form of determinism, we have these premises:

a) What we choose to do is a product of who we are as persons
b) Who we are is a result of prior psychosocial determinants which acted on our minds
c) We had no choice about being exposed to these prior psychosocial determinants
d) Incompatibilism is true.

What we desire, what we *can* desire, what we aim for, and what we *can* aim for, all come from our immediate perceivable environment, or genetics. A newly born child has no choice about its parents or family, wealth, cultural background, race, its own body shape, class, language, religion, socialisation, exposure to peer pressure and propaganda, genetic inheritance, eye colour, physical abilities, mental abilities, etc. Such factors don't merely delimit or circumscribe what a person can think about and understand; if determinism is true, these factors fully necessitate what is chosen by a person. If these factors did not necessitate what a person chose, there would be no explanation for what a person chose. A person's character or self is produced entirely by the

[38] Though the next chapter, on Libet, will be an example of a neurological hard determinism.

[39] Locke, Chapter XXI, Of Power, §25.

above-mentioned factors, for, as we have already seen, the self cannot create itself out of nothing. And it is the person's character that makes the choices.

Certainly, we do think that we can cheer ourselves up, or render ourselves unhappy, etc. I believe that we are under the impression that we can select or control our moods and states of mind. This is not, however, proof that we can "choose" to do otherwise, or be otherwise, or whatever. For all mental states and moods are products of prior mental states or moods, desires, beliefs, etc. Ultimately, those mental states were caused by external experiencings or genetics, and therefore were not up to us to choose. Thus, a hard determinist need not deny the possibility of self-control, but a hard determinist must deny the possibility of ultimate responsibility. And with ultimate responsibility, goes moral responsibility.

Problems for hard determinism

There are four potential issues for the hard determinist position.

Firstly, we may need a stronger defence of incompatibilism. It is not obviously true.

Secondly, we may need a strong psychological materialism—that the mental is fully caused by the physical (psycho-physical identity)—in order to defend hard determinism. For I believe that a robust link between the physical, neurological layer and the mental layer (the "will") is necessary if we wish to prove that some form of *neurological* hard determinism is correct. Some writers, eg., D. Lewis, argue that we might encounter creatures with neural systems totally unlike ours, whom we would not want to deny having some form of consciousness (in Block, 1980, p216 et seq.)[40] This may suggest that we could not defend the view that the mental is reducible to neuro-chemical brain events, and that may entail that hard determinism, predicated on psycho-physical identity, may be false. Thus we may have to defend some kind of strong materialism in order to be able to argue that the we are not free because we are neurochemically determined. Honderich is skeptical of this requirement.[41] (1973, p189). Further exploration of this matter will be beyond the scope of my work.

The third limitation of hard determinism is around Quantum Mechanics. If quantum indeterminism was inherent in nature, then determinism would be false. And some scientists believe this. This threatens hard determinism, because it threatens determinism. But if quantum mechanics was in fact inherent in nature, this would also strip us of the ability to explain many things. Let me elaborate. If knowledge requires that we come to have certain beliefs (always) in the presence of certain facts, we can only know those facts *if* we always come to the same belief

[40] Or consider Nagel's famous example of the bat (1974, p159 et seq.).

[41] Reductive materialism is the view that the mind is only the activities of the brain. Eg., characterisation given in Kim, J. (1989), pp31 et seq.

in the presence of the same facts. Probabilistic causation cannot give this to us; only deterministic causation. So without deterministic causation, we'd only have *probabilistic* beliefs rather than knowledge. Quantum indeterminism would strip us of all full explanations and knowledge.

Finally, there is a worry about consequences. If hard determinism is the case, we may never be responsible for what we do, because prior causes alone would explain our actions, *not* free choices of our *selves*. Morality might fall away completely. I will take this issue up again in the chapter on whether free-will matters.

Hard Incompatibilism

At this point it is worth mentioning Pereboom's *hard incompatibilism* (Pereboom, p89). Recall that this is the view that that since all actions are random, partially random, or determined, there are no free actions. If incompatibilism is true, then free-will is incompatible with determinism. But consider, we have also argued that random (and by inference, partially random) events take control away from the agent, and vest the control in the random event. Yet we have argued that a person needs self-control, at least, to have free choice. Thus, it should be clear, that both randomness and determinism would rob us of free-will. Thus it would follow that hard incompatibilism is true if the premises are true. To put the case another way: If you are determined, what you do is caused by the past which is beyond your control. If your actions are quantum-random, your actions are caused by random events which are beyond your control. Therefore whether your actions are random or determined, they are not within your control. Therefore you have no free-will (Pereboom, p129).

CHAPTER 4
The Timing Experiments of Libet and Grey Walter

Benjamin Libet (1985) and Grey Walter (1993, in Dennett) have conducted neurological experiments which provide evidence that the causes of our actions are non-conscious brain events which are beyond our conscious awareness. But normally, we assume that it is our conscious choices that lead us to do the things we do. If these researchers have correctly interpreted their evidence, it may be that we lack free-will, for we could not control a non-conscious brain state. Libet however has provided evidence that agents can "change their minds" just before performing some action. He felt that this was the elbow-room for free-will. But it may be inconsistent for him to suggest this, since his evidence indicates that there is no room for conscious choices. In this chapter we discuss these results and various objections to the interpretation of the work.

Section 1—Introduction

Most persons, when asked about their choices and their will, would remain in the realm of speech about mental states or conscious states as subjectively experienced, and would make no reference to underlying brain processes. Yet many people today also accept the scientific view that our mental processes, such as the processes of choice, are phenomena inextricably related to or caused by our underlying brain processes.

Benjamin Libet and W. Grey Walter performed experiments to establish the timing of conscious volition. They both found that certain non-conscious brain events significantly preceded conscious intention. The apparent implication of their findings is that antecedent non-conscious brain events cause our actions, rather than conscious "choices". If this evidence were correct, it would entail that free-will does not exist, since we could not control a non-conscious brain event. We generally believe that free choices are made consciously. If our conscious choices were not the cause of our actions, we would not be free, since being free presumably at least means that we consciously choose what we wish to do.

This evidence in itself does not pose a threat to the possibility of free-will without further interpretation, however. Some researchers who argue for the existence of free-will nonetheless take it for granted that some form of neurological determinism is true. But what makes the work we shall examine here interesting, is that the evidence we have seems to indicate that brain processes substantially *precede* conscious mental activity. In other words, if free choice is necessarily conscious, then we could not be choosing our actions (in the traditional sense) if our brain processes preceded our mental processes; our mental processes might turn out to be merely epiphenomenal[42] if they lacked a causal role.

[42] "Epiphenomenal" refers to something caused, but which has no causal impact itself (Dennett, 1993, p402).

Section 2—The Experiments

Libet

In a variety of articles, Libet has provided evidence that a change in the voltage in the brain, called a "readiness potential" (RP)—occurs before intentions do, and before actions do.[43] He reported that the onset of the RP[44] precedes action by about 0.8 seconds (Libet, 1985, p529). Given this relatively long duration, Libet was interested in establishing where the conscious intention occurred in time. We normally think that the intention is the immediate cause of the action, and thus, presumably, conscious intention should also occur 0.8 seconds before the action. But, from our own introspection, this doesn't seem to be the case; we are familiar with the experience of our volition seeming to occur very shortly before our action, not as long as almost one second before.

The subjects in Libet's experiments were asked to decide, arbitrarily, when to flex their wrists, while watching a moving spot on a cathode ray tube. They were asked to report the position of the dot on the tube (after flexing their wrists)—ie., where the dot was at the time they experienced the intention to act. Libet found that the experimental subjects consistently reported that the timing of their inclination to move occurred 0.2 seconds before they actually flexed their wrists. This is what we would expect, from our own introspective experience: our decisions to act *do* seem to occur very shortly before our physical actions. Libet found, however, that there was a significant increase in the readiness potential (RP) approximately 0.55 seconds before the act of wrist-flexing. In other words, the brain was preparing to move the wrist 0.35 seconds *before* the subject of the experiment had "decided" to move his or her wrist! The conclusion Libet drew was that this may threaten the concept of free-will, since we would expect the RP to start at the same time as the intention.

Näätänen (1985, p549) reports the same results for his own experiments. He says he tried to "fool the RP generator" in the brain by first concentrating on the task of reading, and then suddenly doing something else. Näätänen expected to see only a short RP increase before the new act. However, he saw an RP of the usual duration, even though he had thought the act was very spontaneous. In other words, even apparently spontaneous actions, without any preparatory thinking, produced a lengthy RP stage—ie., were not really spontaneous. Libet summarises these findings as follows:

> "Voluntary acts are preceded by ... RPs. ... With spontaneous acts involving no preplanning, the main negative RP shift begins at about -550 ms. ... The time of conscious intention to act was obtained from

[43] The best presentation occurs in Libet, 1985. See also *Behavioural and Brain Sciences*, 1987, 1989, 1990.

[44] discovered by Deecke et al.

> the subject's recall of the spatial clock position of a revolving spot at the time of his *initial awareness of wanting to move* (W). W occurred at about -200 ms. Control experiments, in which a skin stimulus was timed (S), helped evaluate each subject's error ... RP onset *preceded* the uncorrected Ws by about 350 ms and the Ws corrected for S by about 400 ms ... It was concluded that cerebral initiation of a spontaneous voluntary act begins unconsciously [*sic*].[45] However, it was found that the final decision to act could still be consciously controlled during the 150 ms or so remaining after the specific conscious intention appears." [My Italics] (Libet, p529).[46]

If a non-conscious brain event was the true cause of an action, and the decision to act came later in time, then it is apparent that it is not our "decisions" which make us do what we do—or at least—not entirely, but rather there is some non-conscious brain event which plays a necessary role. This sequence of events is clearest depicted on a time-line, represented in Fig. 4.1 below.[47]

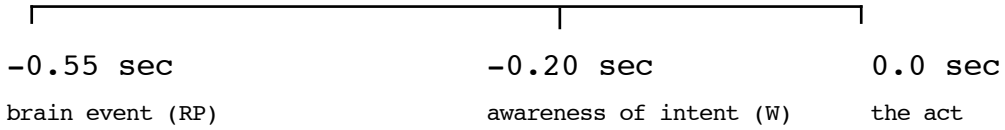


Fig. 4.1—Libet's Findings—1

Libet's concern, therefore, was that W, the feeling of wanting to move, might play no role in the formation of an act; he was worried that it may just be epiphenomenal, or not even play an intermediary role between the choice and the act. Libet is showing that there is some non-conscious brain event (RP) which precedes our awareness of wanting to move. These results, if correct, might strip us of free-will because we assume that the will alone (with proper associated functioning and contextual circumstances) should be enough to cause us to do something (see eg., Honderich, 1973, p197). We assume that it is the *will*, *qua* mental, which determines our actions.

[45] It is important to note that throughout his article and the responses to it, Libet and others use the term "unconscious" to mean "non-conscious". "Unconscious" is usually taken to refer to a mental event of a Freudian type, which could at least in principle be accessible to consciousness (eg. through therapy). Libet's "unconscious" events are however in fact *non*-conscious—they are not accessible to consciousness at all. All quotations in which the word "unconscious" appears, therefore, should be read in the light of this interpretation. It is clear when reading Libet that he means "non-conscious", not Freudian-unconscious.

[46] Libet actually found two types of RP: a type I and a type II. The type-I RP appeared before consciously pre-planned actions, and lasted up to about a second before such actions. The type-II appeared about half a second before *spontaneous* but nonetheless deliberate actions. This implies that in the case of pre-planned actions, we have an even longer prior non-conscious preparatory stage (1987, p785, and 1985, p531, and 1990, p672). This paper concentrates on the type-II RP events since these were the bulk of the types of events that Libet investigated. Thus all actions, whether deliberated and slowly decided on or not, have an antecedent RP.

[47] Note that the diagrams are a substantial simplification of Libet's result graphs, and my diagrams should not be read to entail that the events depicted are not connected by any other events, ie., that they are discrete. Note also that the timing of zero time (the act) was measured accurately by an electromyogram attached to the appropriate muscle.

Libet did, however, find that subjects could "veto" a decision. That is, the subjects were told to sometimes change their minds and cancel their decisions. They often reported being able to do so, and their reports indicated that a conscious decision to *not* do the act occured in the last 0.15 seconds. Libet seemed to think that this was the elbow room for free-will. Libet's evidence indicates that our decisions are more like choices between doing something or refraining from doing that same thing. Decisions, if Libet's evidence is correct, are more like *preventative control mechanisms*—"vetoes" (1985, p529), rather than "triggers" (*ibid.,* p538).

Grey Walter

W. Grey Walter independently performed a similar experiment[48] which differed from Libet's work primarily in that it did not attempt to investigate "veto" possibilities. The subjects (who were brain surgery patients) were asked to press a button to "change" a viewing-slide at any arbitrary time that they desired. The button, however, was a dummy, and the slide was actually changed by an amplification of the RP signal from their brains. The subjects consistently reported that it was as if the slide projector had a kind of bizarre precognition, because it would change the slide before they had decided to change the slide. The subjects reported feeling unnerved that they might accidentally advance the slide twice.

This again illustrates that evidently, the real cause of the act is some non-conscious brain event (RP, possibly), and that it occurs in time *before* the subjective, first-person *apparent* decision event. Our "decisions", as we think of them, may be merely mental epiphenomena, caused by an antecedent neurological event, which is non-conscious, and which in itself represents the actual determinant of the decision event.

Summary

Both the experimental results of Libet and Grey Walter seem to point to this conclusion: that what we call "decisions" or "intentions" are actually mental states caused by non-conscious neural events—and that the conscious aspect of decision-making may be a mere side-effect or epiphenomenal feature of making decisions, without a causal role in our actions. Contrary to our intuitions, the experiments seem to show that it is not our conscious choice which is the sufficient cause of our actions.

> "This [work] seems to show that your consciousness lags behind the brain processes that actually control your body. Many find this an unsettling and even depressing prospect, for it seems to rule out a real (as opposed to illusory) 'executive role' for 'the conscious self'" (Dennett, 1993, p163).

[48] 1963, Presentation to the Osler Society, Oxford University, in Dennett, *Consciousness Explained,* 1993, p167, also in Dennett and Kinsbourne, 1992, p199

We normally think of our choices this way—conscious volition causes action directly:

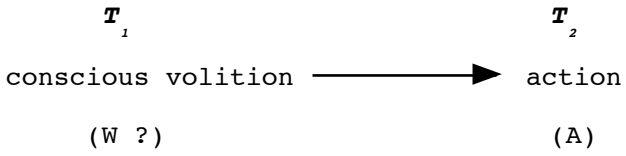


Fig. 4.2—What we believe about choice

If we refer to Figure 4.1, however, we see that Libet's results indicate that *this* is more like what is actually the case:

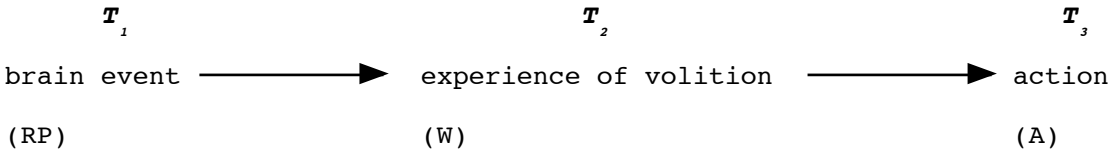


Fig. 4.3—Libet's Findings—2

But if Libet has found something to genuinely be concerned about, then the facts of the matter must look something more like *this*—in which the experiencing of volition plays *no* causal role:

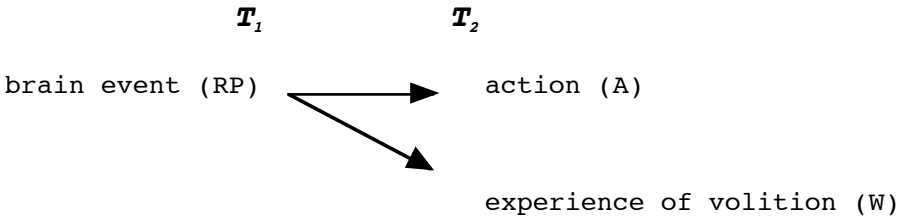


Fig. 4.4—The real concern

Let me interpret this diagram to make it clear. The RP causes the experiencing of volition, and it causes the action, but the experiencing of volition (W) and the action (A) have no causal role in relation to each other. They are independent effects of RP. Libet did also discover, unsurprisingly, that agents could change their minds. He called this a "veto" case. Veto cases, in which the agent changes her mind, do however indicate that mental states could apparently impact the resulting action (A). If the agent had decided otherwise, the agent would have done otherwise. This is correct. For now, this diagram above is merely an expression of what Libet is concerned about; in other words, it is not that Libet is asserting that this *is* the case (Figure 4.4

above)—but his results might indicate that (Fig. 4.4) is correct, and *if* that is what his results indicate, then this represents a threat to free-will, *and*, in my opinion—to Libet's interpretation of his veto cases. But this is an area of contention which I will discuss later.

There are some arguments which indicate that this process, as indicated in Figure 4.4, may be the case, and we will explore these arguments in more detail as we procede. But let us give a preliminary indication as to *why* this process may be correct. Firstly, Libet points out, no special spike in the RP graph seems to correlate with the conscious intention timing (1985, p535). In other words, on his readings, W did not seem to appear. This could mean it was *part* of the readings he was acquiring, or it somehow supervened on the electrical signal he was reading. In any of these cases, there are doubts as to whether W would have a causal role—because the experimental subjects reported W as occurring after RP. Secondly, *if* the conscious experiencing of volition—W—is to play a role, we may have reason to expect it to at least be simultaneous with the RP. But it isn't; it occurs after RP every time, which suggests that it at most is an effect of the RP, not a supervenient concomitant.[49] Thirdly, there is a concern that if W and RP both play a role, that this may involve a case of overdetermination of the action. More will be discussed on these problems later.

Assuming, however, that Libet's evidence is correct, why then would we persist in the belief that it is our conscious selves controlling our bodies? We have two possible reasons for persisting in this belief:

The first reason is this. Are we to attribute the choice of action, or our motion, to the RP, or something even prior to that? The RP occurs in a context where the person, presumably, has already done some pre-planning about what he or she wishes to do, perhaps several days or hours before. So the conscious choice may yet precede the RP. I wish to postpone this critical discussion to a later section.

The second reason must have something to do with the timing. Given that a neural transmission from the brain can take about 0.175 sec (Dennett, 1993, p103) to reach the muscles,[50] and given that Libet says that awareness (W) is reported as occurring approximately 0.2 sec before the act, it is unsurprising that we would think that the consciousness was the cause of the action—because the consciousness and the motor instruction (M) are almost simultaneous. However, for consciousness (W) to causally contribute to our actions, it should clearly *precede* the motor instruction (M); yet it does not. Our illusion of self-control must come from this timing similarity. Whereas, in fact, the two events are roughly simultaneous. To represent this graphically:

---

[49] These are Libet's views and represent his concerns with the implications of his own results.

[50] Starting and stopping a stopwatch. For a speech act it is worse at 0.2 sec.

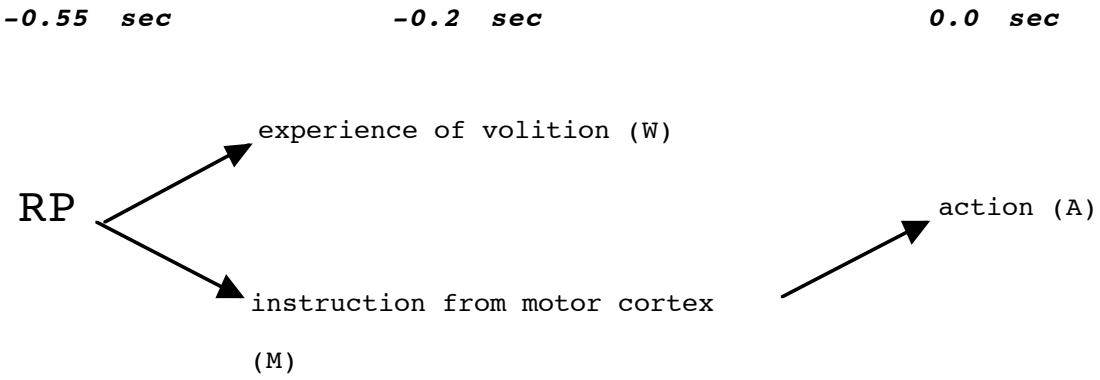


Fig. 4.5—What seems to be happening

Libet reports that subjects could accurately report the timing of the instruction to actually move (M) (1985, p535). This might be taken to indicate that the consciousness caused the instruction to move, but Libet's question then remains as to why a large prior non-conscious RP?

Note that Libet *did* discover that agents could change their minds in the last moments and thereby cancel an action. I wish to leave off discussion of this matter to the section which criticises it, as it will distract us from his other claims, which we now consider critically.[51]

Section 3—Criticisms and Basic Problems

The first question which may arise is what the significance of Libet's work may be. A compatibilist in respect to the free-will debate will readily acknowledge the role of antecedent causation in determining our choices, and yet such a theorist would not feel that this poses a threat to our having free-will. For consider: on most compatibilist models, an agent is free if she can get what she wants because she wanted it, and what she wanted to get was in her best interests, or in accordance with her best reasons. Libet's work is not opposing such a model.[52] Even if an agent's actions are antecedently determined by a neural event, the agent could still be said to be getting what she wants because she wanted it. The question, rather, is one of the causal efficacy of those wants and desires. If the wants and desires were necessary for the carrying-out of the action, then the agent could still have free-will. If, however, the neurological event alone was enough, and the mental events were not causally efficiacious *qua* mental, then they could be irrelevant to the eventuation of the action choice. When we talk of "an agent getting what she wants", the question is one of whether we are talking about the causal efficacy of the RP, or the causal efficacy of the mental state *qua* mental, or both. It is the answer to this which will

[51] Grey Walter is only published in Dennett, 1992 and 1993 (Dennett, personal communication, 2001), thus we shall focus our attention on Libet.

[52] Unless his work requires that the mental is epiphenomenal.

influence our final assessment of Libet's work. This will be discussed in more detail later on. The most pertinent issues of contention around Libet's work can be grouped under a number of different kinds, specifically:

a. The Reports are Subjective—Timing Problems
b. Dennett's timing problems
c. Libet's stance on the Mind-Brain Debate
d. The Plausibility of an Epiphenomenal Account
e. "Act now" vs Reasons for Acting and Preplanned choices
f. Agent Priming and Automation
g. Quantum mechanics explains timing anomalies

and finally, problems with Libet's "Veto"

Some of these criticisms are directly related, so they will appear in a relevant order and be dealt with together. The section on "veto" is quite substantial so it has a section of its own. We shall now proceed with considering these criticisms.

3.a. The Reports are Subjective—Timing Problems

There are generally two types of accusation against Libet on the matter of timing. Firstly, subjectivity, and secondly, that he is ignoring his own work in which he did other timing experiments.

According to this first argument type, the experimental subjects could not have accurately noted, each time, exactly where the dot on the face of the cathode ray tube was—because there were too many things to pay attention to (Breitmeyer, 1985, p539, see also Latto, 1985, p543, Salter, 1989, p181). The argument claims that subjects may have been mistaken about when W occurred, and the RP may have been simultaneous with W, rather than before. The task was too complicated for them to accurately perform it. Either the subject would be paying attention to flexing his wrist (M), or paying attention to the dot on the cathode ray tube, or paying attention to introspecting about when he made the decision (W), or whether a skin stimulus was present (S), or whether he should veto W, or whatever. There were far too many variables to take into account to be able to accurately confirm that the RP definitely occurred 0.3 seconds before the subjective decision to act (Salter, 1989, p181-2).

> "Libet gathered data on two time series: the objective series, which includes ... [an] external clock and the salient neural events ... and [a] subjective series, ... which consists of mental imagery, memories of any preplanning, and a single benchmark datum for each trial: a simultaneity judgment of the form *my conscious intention (W) began simulaneously with the clock spot in position P*" [Dennett's italics] (Dennett, 1993, p163).

Libet, however, was keenly aware of the difficulty of comparing the timescale in consciousness with empirical time (1985, p534), and acknowledged that such reports are subjective in nature (1985, p532, 534). So Libet tried to address such timing problems by using control experiments involving stimulating the skin at the same time the action should occur ("S"). The idea is that if the subjects can accurately report the timing of S (subjectively), they can report W's time, even though it is subjectively ascertained (1985, p534). Libet found that the subjects themselves did not express difficulty doing this. He also found that there was a 50ms difference between W and S, on average.[53] In other words, the experimental subjects were able to give accurate times for S. So if they could tell accurately when they received a skin stimulus, by judging introspectively, they should also be able to tell, introspectively, when they intended to act (W). Especially since these two timing values did not differ significantly on the time scale.[54]

But some critics argued that attending to W is not the same as attending to S—and this would also confuse the timing report (Breitmeyer, Latto, 1985). Furthermore, if Rollman (1985, p551) is correct in his citing a time of 400ms to read a clock position, the subjects would have reported the clock positioning postponed 400ms after they were aware of W. Since we are not sure how much time is wasted interpreting the clock position, we cannot, within a 400ms margin of error, determine the time awareness actually occurred. Wasserman (1985, p557) makes the opposite claim. If the mental processing of the clock takes as long as forming an intention to move, the time reported for the clock position may actually shrink Libet's results, and put the consciousness closer to the start of the RP.

This leads into the second criticism about timing matters, where Libet is accused of ignoring his own previous work. Salter (1989, pp181-2), points out that Libet himself has experimental evidence that there is a large delay in conscious timing perception (p182, see also Dennett, 1993, p153 et seq.). Libet performed other experiments (Libet, 1982) which indicated that subjects perceive stimuli—eg., on their hands—*sooner* than events which directly stimulated their cerebral cortices (Dennett, 1993, p155). We would naturally expect the reverse to be the case. Libet's explanation is what he calls "backward referral"; that the brain puts a time-stamp on the cortical stimulus as having occurred backward in time, as if it had come from an appendage or other more remote part of the body. Libet suggests, in other words, that the brain "edits" its

[53] In Dennett et al., 1992, p184, it is mentioned that there is a 50ms margin of error in timing conscious events. The point is that in Libet, 1985 (*Abstract*), the margin of error is admitted—in agreement with Dennett's data—to also be 50ms or so—and that is not a sufficient error margin to refute Libet's results. Libet's timing anomalies were of the 500ms order of magnitude—ten times larger than the timing margin of error reported by Dennett et al. (1992, p184).

[54] Libet also toyed briefly with the idea that there may be a non-recallable phase of conscious choice starting earlier, but rejected it as non-testable. Another possible explanation he rejected is perhaps RPs just appear, and consciousness still ultimately controls their flowering into action. He rejects this as *ad hoc*, saying it would require RPs to magically appear all the time, at the right time, just when they were needed. Amusingly enough, this is the gist of the enthusiastic response from dualist John Eccles (1985, p543), who says that this *is* what happens: The consciousness "takes advantage" of passing crests of activity, "opportunistically" seizing "the most effective occasions for initiating voluntary actions".

input information before "presenting" it to consciousness (Dennett, 1993, p158). Since it takes a while for information to travel the nerves up to the brain and perception, the brain is apparently set up to "backward refer" stimulus information. Now, if the brain does indeed "refer" perceptual events such as S backward in time, there is no proof that the report of "W" is itself not backward-referred, placing it closer to the start of the RP. Both W and S may therefore be backward-referred in time, because the brain has to compensate for the delays in nerve conduction. Similarly, the report given by the subjects as to when they saw the dot on the cathode tube, may also be backward-referred (Latto, 1985, p545). Salter points out that since Libet et al. found that there is a 500ms delay in achieving what Libet calls "neuronal adequacy"—his term for conscious awareness—that this means there could be inaccuracies in timing W.

Libet's response to the possible backwards-referral or awareness delays, is as follows. He cites (1985, p559), the example of a runner, at the starting blocks, who can start running between 50 and 100ms after the gun has actually fired—ie., almost instantaneously, and a lot quicker than the RP takes to reach "neuronal adequacy". This is surprisingly fast, considering that neural impulses typically take of the order of 200ms. The implication is that neural functions (such as wrist flexing), can have their effects a lot quicker than one may expect, certainly quick enough for the timing of wrist-flexing (as compared to seeing a moving spot) to be reasonably accurate. Furthermore, if sensory processing within the brain was actually as slow as Libet's critics have thus far made out, it could take a runner as much as 500ms to take off.[55] But it does not take that long; therefore, the sensory processing involved, in observing a spot on a cathode ray tube, as well as in waiting for a gunshot, is much faster to process than Libet's critics suggest.

On p560 (1985), Libet makes it clear that the subjects' reports of S and W timings were not statistically significantly different (of the order of 50ms) (see also 1989, p183), and S was at least objective. The difference in timing between S and W may be a reflection of such information-processing delays (that the clock-spot in the W case is a bit harder to process). Libet says that it is consistent with his results that W appears about 350ms after the brain event, as his prior experiments show that stimuli have to persist for about 500ms before they achieve "neuronal adequacy" (ie., awareness, 1989, pp182-3). This may explain why it takes the subject 350ms to realise that an RP event or non-conscious decision had event occured (in other words, W may just be the achievement of neuronal adequacy of RP). What these delays *do not* do, however, is bring the RP within the grasp of *conscious* control (1989, p183). Libet does admit that the timing of the consciousness is potentially problematic (1989, p184) because of the possibility of backwards referral in time (Libet, 1982). And if this were the case, then the report of the timing of W would not be accurate. But Libet says there is no evidence for backward-referral of W, since it is not a sensory stimulus, rather, it is a cumulative effect of the RP. Libet says that he *only* found backwards-referral in time to occur with *sensory stimuli*. Libet says: *that in the volition cases:*

[55] Admittedly, a "priming" argument would dispense with this response. See the section on agent priming.

> "There is no reason to believe such a referral is occurring" (Libet, 1989, p184).[56]

Furthermore, consider that in Libet's sensory experiments (1982), he found that direct cortical stimuli were "backward-referred", as we have seen. If in the case of the experiments we are considering here, there was a backwards-referral, there would be a possibility that W could have been backward-referred. But if W was backward-referred, it would mean that the experiencing of W, like the experiencing of S, had been antedated by the mind as having occurred earlier than it in fact *did*, meaning that W was in fact *later* in time than the experiment subjects reported to occur. Which means that if W *was* backward-referred, it would have occurred in reality, *later* in time than the experimental subjects reported it, making the gap between W and RP *larger*.

3.b. Dennett's Timing Problems

Dennett uses the evidence from Grey Walter and Libet to serve as examples of bad thinking about consciousness (Dennett, 1993, p162 et seq.). He argues that consciousness of perceptual input, such as the position of a spot on a cathode ray tube, is located in a different place in the brain to where consciousness about volitions is located. So we are not entitled to ask "at what time" the agent became conscious of some event *and* some other event in some other part of the brain, since the two parts of the brain function more-or-less independently of each other, and we have no evidence to suggest that the information "comes together" at any point. Libet is thus not entitled to assume that the subjects' reports of the timing of "arriving in consciousness" are accurate, because the two different mental events are not "screened" in one "theatre". The timing of the events need not have occured in the order that they seemed to occur.

> "There is no one inside, looking at the wide-screen show displayed over the whole cortex ... What matters ... is not the temporal properties of the representings, but the temporal properties *represented*, something determined by how they are 'taken' by subsequent processes in the brain."
> (Dennett, 1993, p166).

Vanderwolf (1985, p555), has a similar and strongly behaviouristic view. He cites the example of numbing an experimental subject's arm with an inflatable cuff. The subject, strangely enough, is able to move his fingers at will, but is convinced, if he is unable to *see* the fingers moving, that he is *not* moving his fingers. In other words, our internal states are not as accessible to introspection as we may think they are; we may infer our internal states by observing our bodily motions. This means, contrary to Libet, that if we cannot tell whether our fingers are moving once they are numbed, we cannot also accurately say when W occurred. If this were true, Libet's accuracy of his timing of W may be open to question. But even if Vanderwolf was right, his evidence here, besides bringing Libet's evidence into question, would also serve to confirm that we can act non-consciously. And that's the part we may be concerned about.

---

[56] For a more thorough examination of this debate, see Honderich (1984, 1986).

Commenting then on the Grey Walter experiment, what Dennett says is this:

> "What such a delay would in fact show would be that expectations set up by a decision to change the slide are tuned to expect visual feedback 300 msec later, and to report back with alarm under other conditions... The fact that the alarm *eventually* gets interpreted in the subjective sequence as a perception of misordered events (change before button-push) shows nothing about when in real time the consciousness of the decision to push the button first occurred." (Dennett, 1993, p168, also Dennett and Kinsbourne, p199).[57]

The point, in this case, however, is that the *alarm bell rings*—and that can only mean that something funny *is* actually happening with the timing: the triggering RP *really is* causing the slide to change before consciousness expects to get the feedback from its work, which means the triggering RP really is earlier than consciousness of decision. Even if it is subjective, the evidence still minimally points to a *prior non-conscious brain event* that does the work. Dennett should at least concede that point because he is a materialist (as opposed to a Cartesian Dualist), and *that* is all we need to get the implication of Libet's work through.

3.c. Quantum mechanics explains timing anomalies

There are some arguments to the effect that the observed timing peculiarities are due to quantum time distortions. According to the physicist Heisenberg, the process of measuring the position or velocity of a subatomic particle interferes with the position or velocity of that particle, since we have to use particles to rebound off other particles in order to detect them. Thus, when we try to measure, it engenders an uncertainty in particles' position or momentum (Giancoli, pp751-3). Thus since our minds are chemical systems, they may also be subject to quantum uncertainties.

Stamm has an argument of this form. Self-monitoring by the mind interferes with the other internal processes (such as a volitional act), making for an uncertainty in the timing of its eventuation. Stamm claims that the uncertainty may even be enough to reverse the temporal order.[58] (1985, p554). I find this argument interesting; it might put the control back in the hands of the conscious agent. For if this were true, conscious decisions could be simultaneous with the RP, and then the only question that could arise would be whether the mental event had a causal role. If the mental event is simultaneous with the RP (in fact), it could easily be a causal contributor to the final act, and then Libet's evidence would not necessarily pose a threat to free-will. But if we consider the Grey Walter experiment, the experimental subjects were alarmed that the machine always "knew" in advance when they were going to change the slide. Since the subjects experienced alarm, it does indicate that the RP occurs before the mental event.

[57] This is an example of a "priming" argument—more on this later.

[58] If Stamm were to claim *that*, he may be succumbing to dualism himself, for then the pure mental event would precede the physical event.

3.d. Libet's stance on the Mind-Brain Debate

A very common criticism levelled against Libet amounts to the idea that he is a closet dualist or epiphenomenalist. One critic—Wood (1985, p557)—entitled his response, "Pardon, but your dualism is showing" (see also Merikle and Cheesman, 1985, p548). Let us discuss Libet's evidence to see whether, broadly speaking, he could be accused of subscribing to any such explanatory model. Libet's critics feel that if it is possible to separate mental causation, eg., veto, from neurological causation (RP), that he is implicitly a dualist, since a dualist believes it is possible for mental states to not have corresponding brain states, and yet be causally efficacious. Some commentators have felt that Libet's very research question itself entails dualism. For if Libet believes there are two separate events—RP and W—and they are not simultaneous, Libet may implicitly be importing the assumption that mental events are not token-identical with brain events, and this may imply that he subscribes to dualism. Another example is Libet's veto (V), something we discuss in detail further on. Libet denies that the veto has an antecedent RP; which seems to suggest that the veto is causally efficacious *qua* mental without a non-conscious brain state. But he is not necessarily committed to a view that there is *no* neural correlate of V; it could simply be some other type of neural activity that he was not looking for. Thus, Libet is not necessarily a dualist. He might, however, be an *epiphenomenalist*.

Recall, epiphenomenalism is the view that mental states are effects of brain states, but they have no subsequent effects, themselves. Libet is saying that the mental is caused by the neural; the RP appears to cause both the experiencing of a volition (W) and the action. But the experiencing of a volition, in Libet's evidence, apparently, has no subsequent causal effect. Therefore, if the neural event is causally efficacious, and the mental event is not, then Libet may be suggesting that some form of epiphenomenalism is the case. This is why his research seems to pose a problem for free-will, since we do think that the mental is efficacious *qua* mental. We think that when we have a mental event of choosing to do something, that it is that choice that makes us do what we do. But if epiphenomenalism is true, then actually it's the RP that's causing us to do what we do. On the other hand, if Libet is asserting that the veto (V) *can* be causally efficacious, then not all mental states are causally inefficacious. So it is unclear whether Libet holds epiphenomenalism to be true, either.

It is not clear *what* Libet's view on the mind-body debate is, so we will pursue the idea that Libet's work might entail the truth of an epiphenomenalist account, even if he himself seems to not be committing to any particular position.

3.e. The Plausibility of an Epiphenomenal Account

It is important to briefly discuss epiphenomenalism, since it has its critics, and since it seems as if Libet may be sympathetic to such an account. Dennett (1993) has the following to say:

> "Epiphenomena are mere by-products... '*x* is epiphenomenal' means '*x* is an effect but itself has no effects in the physical world whatever'. ... [this] is too strong. Since *x* has no physical effects (according to this definition), no instrument can detect the presence of *x* directly or indirectly; the way the world goes is not modulated in the slightest by the presence or absence of *x*. How then, could there ever be any empirical reason to assert the presence of *x*?" (Dennett, 1993, p402).

Epiphenomenalism may lead us to "embrace out-and-out dualism" (Dennett, 1993, p403), if our mental states have no empirical import. The kind of epiphenomenalism Libet seems to be describing, however, is not necessarily one which is committed to epiphenomena being causally immune. Epiphenomenalism has certain advantages: it may offer a solution to the problem of property dualism (viz., that mental properties are nothing like brain state properties), and it can rescue us from a threat of overdetermination (see eg., Honderich, 1973, p197).[59] For if epiphenomenalism is true, then as Libet seems to be suggesting, the mental states are irrelevant to the causing of our actions. But this does not require that mental states are non-materialistic, in any sense.

> "Qualia [conscious mental state items], on this reading, *are* physical ... they just aren't functional. Any materialist should be happy to admit that this hypothesis is true—if we identify qualia with reactive dispositions, for instance." (Dennett, 1993, p404).

If mental events or phenomena—such as—"what it is like to have a volition", were in fact just "dispositions"—as Dennett is here suggesting, the *will* would then be "the disposition to act". That makes a lot of sense, and is roughly how Libet characterises volitional experiencing. But the point of Libet's results is just that the agents are only aware of such a "disposition to act" much after the brain activity starts. Defending the epiphenomenal account further, however, is beyond the scope of this paper, and I recognise that evidence against epiphenomenalism would count against Libet.

The question, really, is whether Libet's evidence, implying that epiphenomenalism may be true, is a bad thing. Underwood and Niemi say that it is obvious that the neural substrate is required for consciousness, just as a record is a substrate for music (or as water is required for waves). Thus, they feel, it should not be surprising to us to find some kind of epiphenomenalism[60] is the case. It is clear that mental phenomena, which depend *on* the existence of the brain events, could *not* come first. Libet is merely reminding us that mental events require the antecedent physical ones (1985, p554). But does epiphenomenalism render the mental states causally inefficacious (as Libet fears)?

[59] More on this appears later.

[60] Actually, we could also find that supervenience theory explained Libet's observations as well. The interesting finding in Libet is that brain states *precede* mental states, which may imply that epiphenomenalism, not necessarily supervenience, accurately describes what is happening.

3.f. "Act now" vs Reasons for Acting and Preplanned choices

Consider now the question of what counts as *a* choice. Certainly, before arriving at the experimental chair, the subjects of the experiment chose to be there. Yet Libet did not measure that choice, and that choice of the experimental subject, precedes the choices Libet was studying. Perhaps the RP he detected required those antecedent choices? The agent does have an ability, apparently, to make conscious pre-planned choices before participating in any subsequent action. Even if the ultimate final action itself is non-conscious, it does not mean that it did not occur in a context of antecedent conscious planning and free choice of decisions as to what to do. The argument being suggested here is that the free choice may occur *before* the RP. But there are two answers to this.

Firstly, the subjects did not know in advance that they would be flexing their wrists, so, their antecedent decision to take part in the experiment was not a question of choosing to perform the act of wrist flexing, thus RP was the only immediate antecedent cause, not agents' reasoning. Agents may have chosen to join the experiment, but they did not freely choose to flex their wrists, apparently, since that was non-consciously initiated.

Secondly, Libet (2001) explains that there are *two* kinds of choosing: a *chain of reasoning*, and an *"act now"* decision. Compare the situation where you are "deciding" between two options, and you muse, and reason, and make a decision as to what you will do, but you do not immediately do it. Then consider the kind of choosing where you have two choices present, and you immediately act and choose one. These two types of choice may have different explanations, and may be relevant in different ways for explaining what we ultimately do. Libet argues that it is *not* that chain of reasoning which is causally efficacious in moving the body, but rather it is the "act now" event which ultimately causes the act (2001, p61). Libet's argument in defense of this view is simple: We can muse and reason all day, even taking decisions about what we will do, but never actually *act*:

> "Some may view ... free-will as operative only when voluntary acts follow slower conscious deliberation... But ... any volitional choice does not become a voluntary action until the person moves." (1985, pp538-9).

It is only when we experience the "act now" kind of decision that our body actually does something. Therefore, Libet might argue, it is not our reasons which are the immediate cause of our actions, but rather our "act now" decisions. These decisions, as we have seen, are causally initiated by a non-conscious process beginning 350-400ms before an act.[61] Obviously, one can argue that the antecedent reasoning *contributed* towards the eventual act. That has not been denied. Rather, what has been suggested here is that antecedent reasoning is at the very least not

[61] Searle seems to think that motives *must* be conscious, not all writers agree. McGinn, p36, (Section V), for example.

necessary for action, but that the "act now" event—the presumed result of the RP—might be all that's necessary.

Furthermore, giving ourselves time to prepare does still show a non-conscious RP. Subjects' awareness of a decision to "act now" occurs still only later in time, even if the act choice is pre-planned (1985, p184). As Libet has mentioned, there are two types of RP event; RP I and II. His (1985) article focuses on type-II, associated with spontaneous actions. Libet however found that slow, reasoned actions also have an RP: type-I, which lasted about one second before the action (Libet, 1985, p531, 1987, p785, 1990, p672). Which means that even carefully-planned, apparently freely-chosen acts are preceded by non-conscious brain events. However, if we have antecedently made plans, conscious choices, etc., which we now are carrying out non-consciously, this does not mean that the conscious phase has no contribution to make to that final non-conscious action. I cannot think of any convincing reply to this. But I will speculate, and this is *only* a speculation—that *all* conscious states are preceded by non-conscious build-up of brain activities. See Libet, 1982, for examples pertaining to sensory processing. In that article, Libet discusses how it takes about 0.5 seconds before the consciousness is aware of a sensation, and this is not due to neural conduction delays. Just as, in the 1985 experiments, it takes the subjects about 0.5 seconds to realise that a decision has occurred, it takes them, in the 1982 experiments, about 0.5 seconds to reach neuronal adequacy of sensory processing (Libet, 1989, pp182-3). If sensory processing, *and* spontaneous decisions, *and* motor actions, all are preceded by non-conscious neurological activity, I see no reason to not expect to find the same would be true for *reasoning*. On my hypothesis here, then, we'd find that the antecedent reasoning, too, originated in a non-conscious brain event. This hypothesis is open to scientific testing and I will not defend it further here.

Finally, it is worth noting that the majority of our actions, which we assume are freely chosen, are not pre-planned by some decision policy. Think about it; how many of your day-to-day actions are planned? Most of them are spontaneous, and the choices are taken at the spur of the moment, just like Libet's wrist-flexing cases. Granted, there is a context in which we make our decisions, but many of them can't be explained as being due to such policy decisions. All these uninteresting choices we make every few seconds, Näätänen tells us (1985, p549), have a long antecedent non-conscious RP. Which means that spontaneous, apparently free actions, are the *least* free, because they lack pre-planning and are caused non-consciously. It is not hard to imagine, from here, that we could extrapolate this inference to planned actions as well.

3.g. Agent Priming and Automation[62]

There are two related concerns around the extent to which Libet was studying typical actions. The first is a concern with priming, the second is with automation. The accusation, in brief, was

[62] This argument is drawn from suggestions seen by this author in some commentary on Internet, and in Dennett, et al., 1992, p168).

that he was studying a particular special kind of action which has special features not generalisable to our normal actions. We tend to think that both our spontaneous actions, and our planned actions are free, but we may be ready to believe that merely primed or automated actions are not free. Is Libet conflating these different categories of action? Let's see how these accusations hold up.[63]

Agent-priming

In nature, animals need to make decisions or learn to respond to the environment (Baars, 2000, ppA6-7), *before* environmental circumstances requiring that type of decision or response actually arise. For example, a tiger needs to be ever-ready to pounce on prey, even before any prey is in sight, so that it can "instinctively" attack prey as soon as it is recognised as such, without having to first think about it. Libet's RP events may be this kind of preparation. The subject has been told, prior to the experiment, that he will be flexing his wrist and noting timing on the cathode ray tube (effectively the second hand of a clock being projected), so he *primes himself* before he makes any decisions. This priming may be the RP event (or whatever brain activity underlies it).

One author had this to say:

> "Has anyone considered the possibility that the AEP (average evoked potential) occurring in an already experienced test subject at very short latency and well in advance of actual conscious awareness triggers a preprogrammed motor response such that the subject would not in fact be responding to a conscious formulation of the stimulus but to this much more appropriately timed *trigger signal*? It seems to me that this kind of habituated response would eliminate any need for backward referral both in a test situation and in real life situations like hitting a tennis ball or baseball." (Baggot, 2000, pA6).

And another:

> "So the brain must take time to integrate a state of consciousness (if you believe that neural patterns are the key to consciousness). This means it probably does its best to compensate for the inevitable delays. And it does this first by being an anticipation-based system (predicting as much as possible in advance of it happening) and secondly by having a filter of fast, unthinking, habits that can intercept events at a preconscious level, dealing with them as learnt routines (as when we drive cars or climb stairs)." (McCrone, 2000, pA2)

The ramp in RP, then, could be some kind of preparatory phase, like energy being pent up in preparation to be released. Then, when the action occurs, the energy is released. This does match

[63] Note that for the sake of this argument, we will not be paying attention to the problem of the prior section, namely that prior policy decisions or choice contexts may contribute causally to decisions. All we are considering in this section is the extent to which the kinds of actions studied by Libet were similar or applicable to those we take in our day to day lives.

up with what Libet observed on his graphs. And that is how we think of cases like runners at the starting block, or anyone else who is waiting for the right moment to move. According to this argument, what Libet is seeing is this kind of environmental priming—preparation measures to deal with upcoming circumstances. These preparation measures are not required for action, and they are also not necessarily the direct cause of action. The real decision, on this argument, is still handled by the person's conscious choosing. If the subject of the experiment had been caught by surprise, his response would be just have been a decision-making kind, not a decision-making-and-RP kind (because it would lack the RP, which is merely a preparatory build-up). Under this scenario, the brain events would be simultaneous with the mental events, and the subject would have free choice.

But this argument doesn't explain cases where there is no priming, such as Grey Walters' case where the agent could change the viewing slide at any time, without any other pressures. I hypothesise that the RP may indeed represent some kind of priming, but that the priming is present in all cases and originates non-consciously, since this is what the experimental evidence seems to point to.

Automation

Consider the case of thinking out the task of tying your shoelaces, which is something you deliberately decide to do. It is not done with much conscious deliberation (Doty, 1985, p542, Breitmeyer, *ibid.*). But we do not think of tying laces as involuntary or unfree. The question is whether this kind of non-conscious yet chosen action, is typical for most of our actions. Dennett seems to think it *is* typical (1993, pp251-2).

But these kind of simple actions, as in Libet's experiments, are not the kind of actions we are necessarily interested in, because they have no moral significance. We want to have free-will for deliberate, planned choices. The experimental subjects' choices of action were limited to "stereotyped" actions, which were "inconsequential", and hence not typical of the kind of morally significant actions we are interested in (Breitmeyer, 1985, p539, Danto, *ibid.*, p541). But Libet says these types of actions—deliberated ones—have even longer RPs—type-I (Libet, 1987, p785, and 1985, p531, and 1990, p672). Libet says:

> "It might be argued that unconscious [*sic*] initiation applies to the kind of spontaneous but perhaps impulsive voluntary act studied here, but not to acts involving slower conscious deliberation of choices of action. ... Even when a more loosely defined conscious preplanning has appeared a few seconds before a self-initiated act, the usual specific conscious intention to perform the act was consistently reported as having been experienced separately just prior to each act by all subjects... This leads me to propose that the performance of *every* conscious voluntary act is preceded by special unconscious [*sic*] cerebral processes..." (Libet, 1985, p536). [my italics]

Libet has experimental evidence (1985, p562) which shows that it is *automated* actions—such as those studied—that lack substantial RPs, rather than the other way around. That evidence would suggest, counter-intuitively, that it is *automated* actions that have a lesser RP—and that slow, deliberated actions have a long RP beforehand. Hoffman and Kravitz report this to be true for the case of Tourette's Syndrome patients. They state that persons suffering from this syndrome *lack* an RP before they exhibit a tic, whereas when these patients *mimic* their own tics, they do show a normal RP (1987, p783). Again, this confirms that it is deliberate, supposedly "free" actions that *are* preceded by an RP. Libet similarly maintains (p783 et seq.) that automated motor acts, such as the act of writing[64] letters on a page only occurs if one is aware of a specific wish to *start* or initiate such an action. But that kind of reflection as to why we are doing what we are doing, says Libet, only comes *later* (1987, p784). Thus, the actions which have the least RP are the automated actions—yet we think tend to think that automated actions are the least free.

Section 4—Libet's "Veto"

Libet found evidence that subjects could "veto" their decisions at the last moment. They were able to change their minds. Libet felt that this indicated that we do have some room for free-will. Libet's graphs of the RP in the case of an action, then the case of a veto, look something like this[65]:

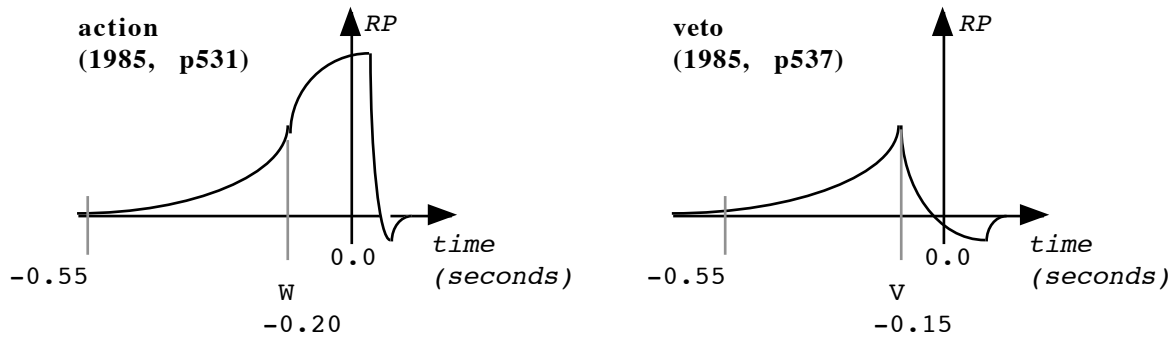


Fig. 4.6—Libet's Graphs of Action and Veto

Libet seemed relieved to have discovered the veto:

> "... the present experimental findings and analysis do not exclude the potential for 'philosophically real' individual responsibility and free-will. Although the volitional process may be initiated by unconscious [*sic*] cerebral activities, conscious control of the actual motor performance of voluntary acts definitely remains possible. The findings should therefore be taken not as being antagonistic to free-will but rather as affecting how free-will might operate ... The concept of conscious veto or blockade of the motor

[64] cf. Jung (1985).

[65] These diagrams greatly simplify the RP measurements which actually appear as a jagged line.

> performance of certain intentions to act is in general accord with certain ... views of ethical behaviour... 'Self control' of the acting out of one's intentions is commonly advocated." (1985, pp538-9).

Libet is also at pains to point out (1987, p783) that although "W" is referred to as "wanting", W is not the actual choice. *Rather, for Libet, choice or free-will comes into effect at the veto stage—subsequent to W* (1987, p784). Choice therefore can only occur in the last 150ms or so of activity (1987, p785). In other words, free-will, as we know it, can at best be a matter of self-control; where we "block" a choice or inclination from following through into action.

Libet also denies that the "vetoes" have antecedent non-conscious neurological causes. On p538, (1985), he says:

> "Would the appearance of a conscious trigger or veto also require its own period of prior neuronal activity, as is postulated for the development of the conscious urge or intention to act and for a conscious sensory experience? Such a requirement would imply that conscious control of the volitional outcome, whether by veto or by an activating trigger, is itself initiated unconsciously [*sic*]. For *control* of the volitional process to be exerted as a *conscious initiative*, **it would indeed seem necessary to postulate that conscious control functions can appear without prior initiation by unconscious** [*sic*] **cerebral processes**, in a context in which conscious awareness of intention to act has already developed" [Italics are Libet's, bold typeface is my emphasis].

There are several problems with Libet's treatment of the veto process, which we now discuss.

4.1. No neural correlate at all, or a simultaneous neural correlate?

Why does Libet think that veto *lacks* antecedent non-conscious neural causes?—He clearly says "it would indeed seem necessary to postulate that conscious control functions can appear without prior initiation by unconscious cerebral processes" (p538). Libet has to take a stance on the mind-body debate (Breitmeyer, 1985, p539). Unless Libet wishes to commit himself to dualism (Nelson, *ibid.*, p550), he has no choice but to assume that the conscious veto is associated with a neurological event *as well*. Whether it be antecedent, like an RP, or whether it be a supervenient relationship with the neural correllate simultaneous with V, there has to be *a* neural correlate (1985, pp538-9).

Does Libet believe that vetoes are not preceded by *any* neurological events at all? If he does mean that, he could be leaning towards the view of dualism. For consider; he seems to be suggesting that V appears *ex nihilo*; lacking a brain state. But perhaps he is not suggesting that. Perhaps, in denying that vetoes have antecedent *non*-conscious neural events, Libet means that vetoes *are* preceded by neurological events, but they are *conscious*. Or perhaps, in saying that V appears without an antecedent RP, he could be suggesting something like supervenience theory; viz., that simultaneous with V is a neurological event which needn't be an RP, which nonetheless

is the neural correlate of V, and which is causally efficacious at halting the choice. Thus, it is unfair, *prima facie*, to accuse Libet of dualism. Perhaps the charge of dualism would stick if Libet was suggesting that V had *no* neural correlate. His writing is not perfectly clear on this matter, however. Perhaps V can be causally efficacious because it supervenes on some simultaneous neurological state, whereas W is not causally efficacious because it succeeds the causally efficacious RP. This is a plausible interpretation of what he means. But this leads us to the next matter of concern.

4.2. Imbalance and inconsistency

Libet, as we have seen, seems to want to allow the vetoes to exist *without* prior RPs so that some kind of mental causal efficacy could prevail (Danto, 1985, p541, MacKay, *ibid.*, p546, Nelson, *ibid.*, p550, Libet, *ibid.,* p538). Libet seems to be suggesting that the veto could do its work, and that the veto is efficacious *qua* mental. But at the same time, Libet is worried that W *requires* a prior RP to be causally efficacious. There is an imbalance here: why is Libet worried that W, a mental event, may *lack* causal efficacy *vis a vis* the eventual action (A), but V, which is also a mental event, apparently *has* causal efficacy? This is inconsistent. Surely either:

> a) W and V are causally necessary for action, and we have free-will because W and V are causally efficacious, and V and W both are supervienient states which have RPs or other neural correlates, and RP and W or RP and V are jointly sufficient and necessary for an action choice. The problem with this option is that it's clear that Libet feels his evidence excludes the causal efficacy of W, but that V is nonetheless efficacious.

or

> b) W is causally efficacious and is part of the causal chain, but it is not free, because it is caused by RP, whereas V is not caused by RP, so it is able to act without the determination of RP. If this were the case then W would be causally efficacious but unfree, whereas V would be causally efficacious and free. The question then would be why V lacks an RP.

or

> c) Neither V nor W are causally efficacious, rather, they are epiphenomenal states having RP or other antecedent neural correlates, and these neural correlates alone are sufficient and necessary for action. This is the interpretation that I favour. I don't think Libet has provided us with direct evidence for this view, but that it ought to be possible to find the evidence.

Since Libet is worried that the RP makes W superfluous, it seems as if he cannot consistently argue that veto (V) is causally efficacious (or if it is, that W is *not*) without explaining why there is a relevant difference, such as the presence or absence of RP. In other words, Libet is worried that RP makes W inefficacious, and postulates that the absence of RP for V makes V efficacious.

There are some problems here. The first question is what *logical* reason Libet has for this claim. Certainly his empirical evidence looks as if it is the case that V has no RP but W does, but it just doesn't seem to make sense. If V is causally efficacious, W must be as well, because they are both decisions and both mental events. Think of it this way. In the case where there is no veto, there is an RP and a W, and the action occurs. This means that either RP or W or both are the cause of the action. If however, there is a veto, V, then neither RP nor W seem efficacious, only V seems to be. Let's assume RP is the cause of action in the non-vetoed case. Why then, would V be causally efficacious if W was not? V is a mental event, just like W. So if V is efficacious, W should be as well. Now if W is causally efficacious and determined by RP, then we would expect that in order for V to be causally efficacious, it would have to be *inside* the causal chain, and must itself be causally determined by something else in the brain; presumably another non-conscious brain state similar to RP. Yet Libet denies this.

Secondly, if the RP or our antecedent choosing is sufficient cause for an action, why do we need the additional veto event to explain it, or to make it free? (Mortensen,1985, p548).

Of course, V could have been determined or caused by some *other* non-RP neurological event. I am not ruling that out. But whether V is determined by an RP or not, V has to be determined by some neurological event. But Libet's belief is very clearly that V *is* free-will, *because* V lacks an RP. It seems as if Libet may be accepting incompatibilist intuitions here: that a neurological cause of V would render V unfree. My concern here is that V and W are both the same kind of thing: W is the wanting-to do something, and V is the not-wanting-to do something. They are both mental, both intentional, and both about an action. Therefore, surely, they must have very similar causal antecedents? This is the inconsistency that I and other commentators on Libet are concerned about. Cetainly, if V is causally efficacious, it exhibits that the mental is causally efficacious. But then, looking back, this would suggest that W would have to be causally efficacious. Yet it is clear that Libet thinks that W is *not* causally efficacious. This is why Libet is being inconsistent.

4.3. That Free-will is more than just changing your mind

It is quite clear from Libet (1985, p539) that he thinks that the capacity to perform a veto is what *constitutes* having free-will (1987, p784). What are we to think then? Suppose W is not free because it is determined antecedently by a non-conscious RP. Then any action caused by W or RP or both, would also not be free. But if V intervenes and cancels W, this means that we are only free when V intervenes. That strikes me as a poor shadow of free-will, because we tend to think we are free when we want to do something *and* we do it; it is not the case that we are only free when we happen to change our minds. Put it another way. Suppose when we only have W and RP, the action is not free because only RP is the cause of the action. This seems to be what Libet believes. Now suppose that under a vetoed case, we are free because V has no antecedent RP. This also seems to be what Libet believes. But this means, if Libet is saying V is free-will,

then all cases where we act but do not change our minds, we are not free, since free-will requires V (changing your mind). That strikes me as odd, to say the least.

A further consideration about why we have to be free to choose to *do* things, not just to *cancel* them, occurs in Doty:

> "If the preparatory movement is wholly outside conscious control, how could a conscious process then 'know' what will ensue if it fails to veto the brain's proposal? In this scheme, consciousness is relegated to an intuitive process of guessing what it may be that 'the brain' is up to..." (1985, p542).

How, in other words, could we possibly know what we were going to choose, in order to veto it, if W is not under our control? How could it be that we could *not* be free when generating W, and yet when generating a change of mind—V—we *are* free?

4.4. Why W and V may both not be causally efficacious

Apart from the inconsistency in Libet, the most substantial reason for rejecting the causal efficacy of V is this. Libet's evidence shows that W occurs at the same time as M (the cortical initiation of the action). Thus, we have no good reason to suppose that W contributes to the eventuation of M. Rather, it looks as if W is merely an effect of RP, just as M is. Analogously, and this is the crux of the matter: since V occurs in time shortly *after* M, we have no reason at all to suppose it is responsible for influencing M; rather, there must be some antecedent neurological event which halted M. Remember, neurological transmission time from the motor cortex is at best 175-200 msec, which gives V or W no time to influence M. *Since V occurs at 150 msec before the action, and since M occurs at 175-200 msec before the action, V cannot halt M from being sent down the arm to the hand, because the signal has already been sent.* Libet reports that "neuronal adequacy" or awareness of neurological events takes around 200 msec to be realised *after* the neurological event has occurred (1982); therefore V and W, *qua* "neuronally adequate", could not influence M in time, because they take 200 msec to reach that state. In other words, while W and V are building up over the 200 msec period to achieve neuronal adequacy, the same RP that caused them *also* initiated the M signal.

If it takes the consciousness about 200 msec to become apprised of a neural event, and if RP is the neural correlate of W, occuring about 300 msec before W, we would expect there to be an antecedent brain state before V. Since V occurs at about -150 msec, we'd expect its neural correlate to occur at about -350 msec to -450 msec. *That* neural event would be far back in time enough to cause M to halt. My suspicion is that were Libet to redo his experiments and measure for veto brain events, he would find something more like Figure 4.7 below.I am hypothesising, for the sake of future experimental testing, that antecedent to the veto decision, V, there will be another brain event, let's call it $RP_{(v)}$, which is the neurological antecedent of V, just as RP (here

called $RP_{(w)}$)—which Libet detected—is the neurological correlate of W.[66]

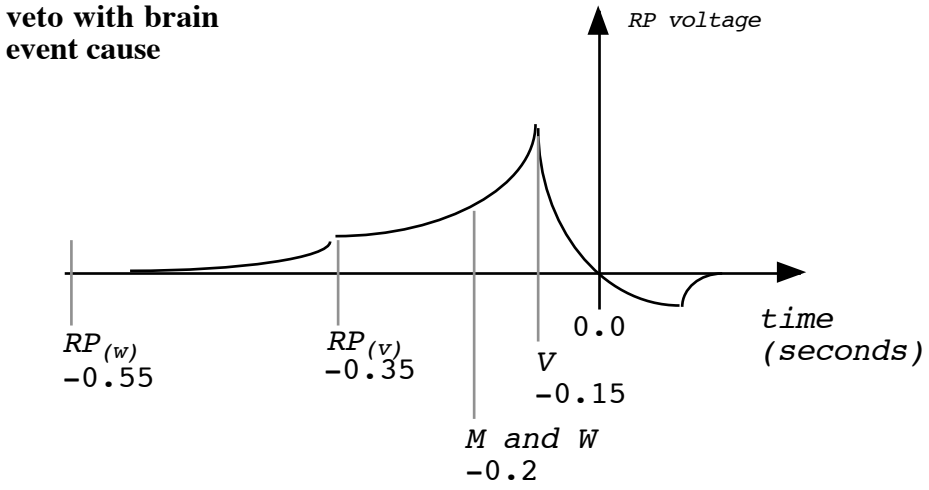


Fig. 4.7—A possible solution

It is possible that under experimentation, Libet might *not* find an RP *per se* for veto, because RPs are typically associated with physical movements (Wood, p558, Libet, p562). However, Libet should at least see *some* neural activity indicating some prior brain events which lead up to the later consciously-experienced veto.

Section 5—Summary and Conclusion

It seems as if Libet is seduced by the idea of having free-will and control over his actions, and despite his own evidence that actions are non-consciously initiated, Libet wishes to privilege "veto" acts as not having antecedent non-conscious physical causes, in order to rescue free-will. To be sure, he could think that vetoes are associated with conscious neurological events, in which case he'd not be guilty of dualism. But then he'd still be guilty of inconsistency. For Libet is concerned about the possibility that W is not causally efficacious: indeed, his whole fear was that the preceding RP indicated that W was just the experiencing of a volition to move. Libet is very careful about this. He never calls "W" the "will"—the decision-making entity. W is always characterised by Libet as *just* the *awareness* of wanting to do something. Libet felt that it was important to distinguish W (the awareness of wanting to do something), from the will *per se*, which is manifested in the later "veto" occurring at -0.15 seconds (Libet, 1987, p784). This is a distinction that Libet makes because he wishes to preserve the power of the will to decide, *qua* mental.

---

[66] In this diagram, $RP_{(v)}$ is the neural event preceding and sufficiently causing the veto, and $RP_{(w)}$ is the neural event preceding and causing the act M (unless a veto occurs). I am not committing to the view that there *is* an RP *as such* prior to V (veto), just that there is *some* neural correlate of V. It is called $RP_{(v)}$ for ease of analogy

In the case of V, however, Libet is quite clear that he thinks that it *is* causally efficacious, and that it is a manifestation of free-will, traditionally construed. *Remembering now that Libet thinks that free-will is veto,* we are left with the amazing conclusion that if W is *not* the will proper, then in a non-vetoed case, the *will* may be reduced to a role of passively waiting for a reason to intervene in actions which are generated spontaneously somehow in the non-conscious part of our minds. This sounds odd. Surely we *can* freely and voluntarily initiate actions even if our choice will not involve a veto (changing our minds)?

I believe, however, that V is not the *decision* to veto, but rather just the subject's *awareness of a decision* to veto. Since veto is also a mental event, and since it must be associated with a neural event, then it makes sense to assume that veto is preceded by its own neurological event which could be non-conscious and which causally necessitates the cancellation of the event M. But if we are correct in arguing that W and V are epiphenomenal, we may wonder why we still have them or experience them. I cannot answer this with certainty; one is tempted to think that these states *must* have a purpose, but I cannot think what it may be, because epiphenomenal *means* that they have no effect. But this is entering the mind-body debate, which is beyond the scope of this paper.

The following might be a more plausible interpretation of Libet's results: that the will (as we know it) is not efficacious, and that what we call "decisions" are in fact the epiphenomenal results of non-conscious prior brain events, and that W really is the will—and it lacks causal efficacy. Otherwise, the will would *be* the RP—in which case our choices would be non-conscious. Libet's results *show* that the veto is at least causally efficacious. However, as I argued earlier, the veto itself could *also* just be an epiphenomenal experience. Just as W is an awareness of something having been chosen, so the "veto" could turn out to be the awareness of something having been vetoed.

The only objection to my interpretation of Libet's results which seems to have any promise, is that objection which says that the RP forms in the context of already-chosen policies of behaviour, where the agent has already decided what he or she is going to do, and the RP is just an event that occurs at the time of the act itself; it does not represent the reasoning or choosing *per se*. This argument is hard to disprove, but to remind the reader, I have speculated, in response to this criticism, that *the reasoning itself* would have a similar relationship of epiphenomenalism to its neural substrate. I recognise that the burden of proof here is mine.

We know we can control ourselves in some sense, and veto our decisions. But I believe that what we decide is a *non-conscious* result of our non-conscious self-structures (see eg. Dennett, 1993, p199 et seq.). It is only through the possibility of the existence of a self, therefore, that we could still have free choice. This I investigate in my next chapter.

CHAPTER 5
The Existence, Nature, and Function of the Self

> In order to legitimately ascribe moral responsibility to some human body, we need the body to have a "Self", which causes the body to move, which is said to "make the choices" that the body carries out. What distinguishes mere motions from actions, it is said, is whether a "self" *willed* the motions.[67] We do not apply moral judgments to mere motions, but we do apply moral judgment to *actions*. Thus, without a self, or one which is structured as we believe is required for moral responsibility, people would just be moving rather than acting, and would lack moral responsibility. I aim in this chapter to investigate a variety of models of what the self is, and whether these models provide adequate grounding for our moral responsibility. If we lacked such a self, we might find that we lack moral responsibility, and therefore, lack free-will.

Section 1—What is a self, and why is it needed for moral responsibility?

i. What a self is

Let us begin by consulting some commonplace defintions.

> "**self** ... **1** a person's or thing's own individuality ... **2** a person or thing as the object of introspection or reflexive action (the consciousness of self)." (OED)

This alone, however, doesn't quite capture what we mean by 'self'. It is quite common, for example, to slip between the concept of "self" and concepts like "I" or "ego". We often say things like, "I, myself..."; someone who is "preoccupied with himself" is "egotistic", etc. The self could thus be regarded as the same thing as the "ego" or the "I".

> "**ego** ... **1** *Metaphysics* a conscious thinking subject. **2** *Psychol.* the part of the mind that reacts to reality and has a sense of individuality. ..." (*ibid.*)

> "**I** ... used by a speaker or writer to refer to himself or herself. ... *Metaphysics* the ego; the subject or object of self-consciousness." (*ibid.*)

The word "self" is thus a psychological term, referring to that which has our personality traits and memories. Our "selves" are aware of what our senses take in from the environment. Our inner life is accessible to us, and our sensory input seems to appear to us in a unified, multi-modal display in our "mind's eye", as if we have an inner "stage", or "theatre". The self is some kind of *inner observer*. Searle, in *Consciousness* (1999) describes the nature of

[67] This chapter does not discuss "self creation" as this is mentioned, in passing, in the chapter on libertarianism with reference to Glover and Sartre.

consciousness as "essentially" a "Combination of Qualitativeness, Subjectivity and Unity", and that consciousness is "by definition, unified" (*ibid.*). Even Velleman thinks of the self as a unified "arena" (*sic*, p123) where our decisions and thoughts and visions are played out.

ii. Why we need a Self for moral responsibility

We use the concept of a "self" as a label for what we take to be the cause of our actions (McDermott, 1992, pp217-8). The self thinks, reasons, and decides between our options. The self is the causal originator of the actions of the body. All choices for which we are responsible have to ultimately have their explanations terminate at the self.[68]

Consider this. We do not apply moral judgments to mere motions. In order for some motion to have moral import, it has to reflect the desires of a conscious free agent. Searle (2001a) discusses this matter;[69] he offers us a theory of the Self, as follows. A self, according to Searle, is "conscious agency plus conscious rationality" (p511).[70] We want the explanations of what we choose to do, to provide causally sufficient conditions: that is, conditions, which all taken together, will ensure that only one particular thing is chosen. But Searle expresses some doubts that the causal antecedents of our decisions are sufficient to explain our actions (p495, p499). Only if we make reference to the reasons for actions, he argues, can our explanations for our actions be adequate (*ibid.,* p492). For the difference between a mere motion and an action is that a self wills the action. But for a self to freely choose an action, the self must have a reason *for* that action. In order for those reasons to be *our* reasons, the reasons that we have would have to be the reasons of *a* self. Searle thus concludes (unconvincingly) that we need to posit the existence of an *irreducible self* to give our bodily motions *reasons*, and in so doing, turn them into *actions*:

> "The logical form of the statement 'Agent S performed act A because of reason R' is not of the form 'A caused B', it is of the form 'A self S performed action A, and in the performance of A, S acted on reason R'. The logical form ... of rational explanation is quite different from standard causal explanations. The form of the explanation is not to give causally sufficient conditions, but to cite the reason that the agent acted on. ... But if that is right, then we have a peculiar result. *It seems that rational action explanations require us to postulate the existence of an irreducible self, a rational agent*... [my italics]" (p500).
>
> "... when spelled out, *the logical form of such explanations requires that we postulate an irreducible non-Humean self*. Thus: ... Reason explanations are adequate because they explain why a self acted in a

[68] This is not to pre-judge the metaphysical nature or structure of the self may be; at this point we are merely describing the commonplace characterisation of the self and its relationship to our actions.

[69] See also Libet's (2001) response to Searle.

[70] Debate on reasons-based characterisations of free-will can be found in the chapter on compatibilism.

> certain way. They explain why a rational self ... acted one way rather than another... and for their intelligibility, these explanations require that we recognise that *there must be an entity—a rational agent, a self, or an ego—that acts ... because a Humean bundle of perceptions would not be enough to account for the adequacy of the explanations*" (p501) [My italics].[71]

Searle seems to be starting from the premise that only a self can have reasons. Thus, without a self, which can have reasons, we could not do things *for* reasons. Thus, in order to do things for reasons, we need a self. If we could not do things *for* reasons, we could not act freely, since free acts are those acts in which we act on our reasons—ie., where we carry out certain behaviours in order to fulfil our desires, based on some beliefs we have about how to satisfy those desires.[72] Searle however goes one step further and claims that this model of the self—the irreducible self—is the only one capable of allowing us to have real free choice and preserving our moral responsibility (p502). Searle claims that neither a constructed self, nor a Humean bundle of perceptions, could be adequate. It seems to me as if Searle's argument can best be expounded as follows. Whether this summary below proves Searle's point that an *irreducible* self is required, is not clear.

1. Moral reprobation is justly directed only at one entity: the entity which caused the morally relevant event.

2. To be morally responsible, an entity would have to be apprised of the moral significance, rightness, wrongness, and benefits or harms which would be engendered by the performing of an action which led to a morally relevant event. The entity would have to know the reasons for its actions.

3. To be morally responsible for an event, therefore, an entity would have to be rational.

4. To be morally responsible, Searle argues, an entity has to be irreducible, because we cannot direct moral reprobation at anything other than the *particular* cause of the morally relevant event. Double suggests, similarly, that we must have unity for agency:

> "For each human body, there [must] exist... one agent who is the subject of the mental states associated with that body." (p46, also p47). There must be no "multiple recalcitrant cognitive subsystems" (Double, p38).

I realise that these points are contentious, viz., that the self has to be irreducible, or 'single', or 'unified', in order to be causally responsible. More on this will be discussed in later sections, where I will try show that because we lack such a self, we lack moral responsibility.

[71] By "Humean" here, Searle is referring to Hume's view that the self is merely the collection of our perceptions. Hume's view will be covered in more detail later on.

[72] See also Leon, Wolf, Velleman, et al., on "tracking" models of free-will.

5. Since only irreducible selves can be causally responsible, and since only rational selves can be morally responsible, only irreducible, rational selves can be free and morally responsible. Thus only irreducible selves can be morally responsible (4), because it seems as if only irreducible selves can be rational, as moral responsibility is entailed by rationality (3). This point doesn't directly play a role in his argument, as far as I can tell, but it is one of his conclusions, which is why we mention it.

6. Therefore we need to have irreducible selves to be morally responsible.

7. Therefore we cannot be free if we do not have irreducible selves.

Problems for Searle

Searle's argument interests me because it seems to line up very well with our traditional intuitions about what a self *is*, and how it is involved in the process of voluntary actions, choices, etc. But there are some problems, particularly:

(a) There are a wide variety of arguments which defend some or other model of human free actions, and many of these arguments make use of the idea that an action is free if it is acted upon because the agent in question had a reason to act. But it is not clear that we need a self, as Searle construes it (separate, single, unified, etc.), to make such arguments possible—ie., it may be possible for agents or selves to make choices and act for reasons even if those agents or selves are not "irreducible".

(b) In his papers (2001a, 2001b), as we see in the quotation above, Searle tries to argue that the "self" needs to be "irreducible" in order to adequately explain action—but what he means by "irreducible", is anything but clear. The term "irreducible" suggests to me, a kind of atomism or indivisibility with respect to its metaphysical structure.[73] But it isn't perfectly obvious *why* Searle feels that this particular type of self—an irreducible single unified self—is required. If we think, for example, of how a committee can be said to be responsible for a decision, yet a committee is neither a single thing nor unified, it seems as if a structured entity could be causally responsible. To give the points of view in brief:

- It could be argued that there is no irreducible self and thus there cannot be free-will
- But Searle argues that there is an irreducible self and therefore there is free-will

[73] Searle's usage of this term "irreducible" might not imply the Cartesianism dualism that Dennett attacks, but it might imply a kind of Cartesianism in which the self is a separate physical entity, perhaps "supervening" on the physical layer in some way. Searle in fact *does* argue for something like this in the relevant article (2001a). He describes the self as a free-floating, probably electrostatic field "over" the whole cortex. But that sounds epiphenomenal, and may be open to doubts around its possible causal efficacy.

• Other writers argue there is no irreducible self and yet there can be free-will

Let us now consider these points of view in detail.

Section 2—Models of the Self—Introduction

In order to ascertain what relevance the self has for our moral responsibility, we need to decide which model of the self would give us the kind of robust link we need between our decision-making processes and our final actions. If we can find a model of the self that can provide a strong link between our decidings and our doings, then we could have free-will. Similarly, if we lacked such a self, we could not be morally responsible and would lack free-will.

There are three models of what we call "the self", that will be discussed.

One traditional view is that we have a self—a single unified irreducible entity—that controls our bodies, makes our choices and is responsible for what we do, and has our experiences. We will call this view the *Cartesian model of the self*. This view is however not committed to Cartesian *dualism*.

The next view we will consider will be called the *Constructivist* view. Constructivism describes the self as a system of mental states or mental entities which cause our actions. This view has two threads: The first characterises what a self might be and how it might be causally efficacious, and the second specifies how to discern *when* a motion is an action. Constructivists ask: Under what conditions would we take a self to be acting, or under what conditions would an action reflect a decision of a self? To put it another way: Constructivists are interested in discerning when an action reflects the choice of a self—when an agent is in an action—as well as when a motion is an action. This view argues that in order to be free or morally responsible, there has to be a certain sort of systematic relationship between the desires, values, and reasons of the person. The self, on this model, does not have to be a single unified irreducible entity.

The final view we will consider will be a skeptical view, which will ask whether the self exists at all, in any plausible sense, and whether it could control our bodies if it did exist. The argument of the skeptic is this: There is no particular mental entity which can be ascribed with responsibility for our actions. Therefore, we are not free. This model agrees with the Cartesians on the intuition that we need a single, unified irreducible entity—the self—to take responsibility for our actions, but, it agrees with the constructivists that the self, if anything, is a system. This view then concludes that since the self is not unified, irreducible or singular, it cannot be causally responsible.

Section 3—The Cartesian Model of the Self

The Cartesian model characterises the self as follows. The Self is (a) the primary cause of our actions, (b) the mental arena of our experiencings, (c) unified, irreducible and singular, and (d), the entity which reasons and makes our decisions. The "Cartesian model of the self" is not the same as Cartesian*ism*, which has *dualism* as an additional premise. The Cartesian model of the self is not committed to dualism of mind and body, even if Descartes' work provides an example which is so committed:

> "..so that *"I"*, that is to say, the mind by which I am what I am, is wholly distinct from the body, and is even more easily known than the latter, and is such, that although the latter were not, it would still continue to be all that it is." (*Discourse on Method—Part 4 ¶ 2*)

You may think of "Cartesianism with respect to the self" as meaning what Descartes meant except the self *could be but doesn't have to be* a free-floating soul—*it could be something else*, eg., *a* brain event, or particular neuron, or the activities of a particular neuron, or *a* Searlian Field (2001a, 2001b), or whatever—as long as it is some particular irreducible thing. The argument of the Cartesian theorist would go approximately as follows:

a) It is obvious we are able to make choices that we carry out, and

b) It is apparent that our consciousness is unified and all inner experiencing is accessible to one entity—the person who is having those experiences

c) Since there is one inner entity—the self or person—who chooses the actions of the body, it is evident that that entity is the one which is responsible for what the body does.

d) Only a singular entity, indivisible and in control of the body directly, could be responsible for what the body does. This point is defended by reference to the fact that we are apparently such an inner experiencing being, and that we are morally responsible. Since it is only beings such as us who are responsible—beings with a rational inner unified self—our responsibility, as selves, must originate in our being such entities. It is *we* as individuals who are responsible.

Problems for Cartesianism

The Cartesian model of the self does seem to tie up with our intuitions about how it is that we come to know about the world, and decide what to do *in* the world. But there are some substantial problems with it.

Our first question is thus: if a Cartesian self is required in order to make moral responsibility possible, then we need a candidate to fill that role—and it would have to be a plausible candidate, with suitable structural complexity to provide the complex features that a self has, such as consciousness, the ability to make decisions, a personality, etc. We may find it difficult to decide which candidate suits this role. Descartes thought it was the soul, but it is well-known now what the problem with that theory is; viz., how it interacts with the physical body. Other variants on the Cartesian model of the self may have similar problems, such as the *site* of the interaction with the body. Descartes thought it was the pineal gland; again, we know better these days. The criticism here is that we need a candidate for the role of a self, and a mechanism to explain how it interacts with the body. Certainly all Cartesian models provide such candidates, but as we shall see in the later section on Dennett, these candidates and explanations have problems.

Secondly, let's take Descartes' famous *"Cogito ergo sum"*: Just because thinking exists, it does not prove that there exists a *self* or "I" which is *doing* the thinking. Refer to the above point; we think we are a single inner entity which does everything; the thinking, the choosing, the perceiving, etc. Someone has to exist, presumably, to *have* those thoughts, decisions, etc. Persons, says Parfit, must be included in descriptions of semantic content of thoughts, memories, etc. For those states are not mental states of a person, *unless* their description involves reference to a person. But persons (or selves) do *not* have to be posited as existing as separate entities (Parfit, 1984, pp225-6). The existence of a self, argues Parfit, does not involve some "further fact" over and above the continuity of mental states (p242). This means that it is possible that our choices are not made by a single Cartesian self, but rather our choices are a result of a collective of mental states which lack an overall controlling entity.

Third, we have seen in the previous chapter that Libet provides evidence that the mental is not causally efficacious. If the self is a mental entity, the self cannot be causally efficacious. Thus the self could not be responsible for what our bodies do. If the self were a free-floating entity, such as is posited by Descartes, it too would lack causal efficacy. If we are to explain consciousness at all, without it being some kind of inexplicable brute fact, at some stage it has to be reduced to non-conscious components (Dennett, 1984)—otherwise we will keep on making circular reference to irreducible consciousness elements. If we are to explain what makes something *alive*, at some stage we will have to resort to talk of mechanisms involving only non-living parts. Now consider this *reductio ad absurdum:* If voluntary actions are produced by a self's *volitions*, we need to ask whether those volitions are themselves voluntary. If they are, we get an infinite regress (or a circular definition). If however we recognise that voluntary choices have to eventually be explained by something non-voluntary, to avoid this circularity, we obtain the result that voluntary acts are the result of non-voluntary events (Ryle, in Dennett, 1984, p78). Dennett says:

> "Are decisions voluntary? Or are they things that happen to us? ... Our decision bubbles up to consciousness from we know not where. We do not witness it being *made*; we witness its *arrival*. This can lead to the strange idea that Central Headquarters is not where we, as conscious introspectors, are;

> it is somewhere deeper within us and inaccessible to us. ... Why must there be a center at all? The illusion of such an ultimate center arises, I think, from our taking a good idea, the idea of a self and unitary and cohering point of view on the world, and pushing it too far under the pressure of preoccupations with our [moral] responsibility. ... Faced with our inability to 'see' (by 'introspection') where the center or *source* of our free action is, and loath to abandon our conviction that we really do things (for which we are responsible), we exploit the cognitive ... gaps ... by filling it with a rather magical ... entity, the unmoved mover, the active self" (Dennett, 1984, pp78-80. Italics are Dennett's).

This is not to say that voluntary choices *have to* be non-conscious, but if they are non-voluntary, my bet would be that they are *also* non-conscious, as Libet indicates. Thus even if we had a conscious soul, its conscious deliberate choices would most likely have to ultimately be explained by non-conscious, non-deliberate events.

Lastly, we have the substantial arguments from Dennett and Kinsbourne which directly criticise the cenralisation tendencies of the Cartesian model of the self.

How Dennett dismisses the Central Observing Cartesian Self

In *Behavioural and Brain Sciences* (1992), Dennett and Kinsbourne suggest that our normal concept of our inner mental life, is mistaken. A mind is commonly thought to be a "locus of subjectivity", it always involves a "point of view". It is thought of as "a thing it is like something to be".[74] Dennett says: "It is tempting to suppose that there must be some place in the brain where 'it all comes together' in a multimodal representation or display" (p183). But, he says, it is a 'mistake' to hold this view (p185). Dennett et al. thus deny that we have a single locus of perceptual consciousness.[75] This objection is also raised in Parfit (p249). Dennett et al. highlight two problems with this assumption of centrality: a philosophical problem and one related to evolution and survival.

The philosophical problem is this. According to Dennett (1993, pp52-3), what he calls "Cartesian materialism" (the single-self view) involves an infinite regress. The *homunculus* (Latin: little man) who is looking at the screen or stage "where it all comes together" (Dennett et al., 1992, p184) would need his own screen or stage—Cartesian Theatre—if he is to be conscious. And his *homunculus* also has to have a Cartesian Theatre, etc. Dennett's move, therefore, is to deny the existence of the very first *homunculus* (1993, pp52-3)—and with that goes the Cartesian self. By way of dismissing this concern, however, one can simply deny that a conscious self is anything like a *homunculus*.

[74] a reference to Thomas Nagel—"What it is like to be a bat".

[75] Recall our definition at the beginning where we identified that one of the primary features of the self is that it is the locus of our consciousness; where our perceptions all come together. If the self is a whole entity, a threat to its claim to being a central unified perceptual arena is also a threat to its existence as a whole.

The second problem that Dennett et al. highlight, is this. Grant that we are able to recognise when two events occur simultaneously. In order to ascertain timing simultaneity of experiences, we would presumably need a central locus of consciousness (Dennett et al., 1992, p184). Since we know we can make judgements of simultaneity accurately, it would follow that we have a central locus of consciousness. Let's then do a modest thought experiment. Take the example of a simultaneous tap on the forehead and toe. The toe-brain distance is longer than the forehead-brain distance, *so one may imagine*, that in order to perceive two taps (one on the foot and one on the head) as simultaneous (which they really are, and which one can introspectively *tell* that they are), one would need a "delay circuit" to keep the tap on the forehead out of consciousness until the tap on the foot "arrived". But, Dennett asks, why should sensory input be delayed *just in case* something else relevant comes along later in time? (1992, p188). This would not make evolutionary sense: How would our brains "know" *in advance* to delay the forehead tap and keep it "out of consciousness" *until* a foot-tap arrived, so that the stimuli could arrive literally simultaneously in a single consciousness? This clearly is wrong, since the brain lacks precognition. This would mean that to obtain the impression of simultaneity, we'd need to do some postprocessing. Thus our awareness, in all likelihood, must therefore self-correct over time as more information comes in. This suggests that there really is no "final screening place" which is required for conscious experience.

Thus, Dennett et al. conclude, if there is no central arena of perception, there could be no central self doing the perceiving.

> "The pineal gland is not only not the fax machine to the Soul, it is also not the Oval Office of the brain, and *neither are any other portions of the brain*.". [My italics] (1993, p106).

I take this to mean that there is no central control structure in the brain which is responsible for the decisions we take. Dennett has a number of arguments, and he begins by rejecting centralised apperception. After that, he argues for decentralised self-control. That again leads him to deny that there is "something in charge".

Cartesian materialism is therefore false. Dennett et al. say, (1992, p235): Asking about when "I" (my self) became aware of something is like asking when the British Empire became aware of the *Truce of the War of 1812*—because *"I"*, like the *Empire*, is an *organisation* of many things spread out in space and time. Dennett concludes that the self, if anything, is a system of mental states.[76] If selves are systems spread out in space and time, Dennett suggests, then there is no guarantee that the *system as a whole* is ever apprised of anything at the same time as its parts (1992, p236). And more importantly, it doesn't *need* to be; that wouldn't make good survival sense. It does not make sense to suggest that, for example, information being processed by the auditory centres of the brain must be passed through the visual processing regions of the brain as well, so that our "selves" could have their "TV set" play sound as well as video. Dennett

[76] this claim agrees with Constructivism

does not, however, seem to draw the inference that the same argument applies to responsibility. But the same must apply. Since "I" cannot be precisely said to have been apprised of some fact at a particular time *because* the "I" is fragmented and distributed, we cannot pinpoint causal origination in the way we might want to, because the *supposed* causal originator (the self) is also spread out in space and time. Yet we *do* think of actions as having particular single causes which are morally responsible for those actions.

There are two counter-arguments to Dennett that appear in literature. The first accuses him of attacking a straw man.[77] Many authors, in response to the Dennett and Kinsbourne article, have pointed out that very few philosophers today would defend a Cartesian viewpoint of the self. For if Dennett et al. cannot escape from a reliance on a description of consciousness which involves the arrival of a stimulus from the bowels of our unconscious up into conscious light, then they too are talking of a central Theatre-Self, an "arena" where it all happens.[78] The "Theatre" that Dennett et al. are attacking thus seems to be a straw man. Dennett et al., say that consciousness consists of sequences of states which replace each other as new information arrives. However, the other authors observe, this does not rule out the possibility that there is some internal observational point, or central presentation point, of the material.[79] As Lycan observes, Dennett and Kinsbourne's model is not anathema to an "internal scanner" view; some kind of *"attentional function"*—ie., where we are able to exercise a function of paying differential attention.[80] We could, therefore, even under Dennett's model, be able to pay attention to different aspects of our internal life without committing us to being *a particular thing doing the looking-at*. Moreover, perhaps all that Dennett has done is create many "theatres", in positing many "drafts".[81] All this criticism is however moot, since all of it acknowledges that there shouldn't be *a particular thing doing the looking-at*.

Aronson et al. point out, however, in a second criticism, that Dennett can only get a model of multiple *selves*, if he *proves* that we lack any single stream of consciousness, that there really is no "place" where it "all comes together". Roskies et al. object similarly to Dennett as follows: *Who or what* does the "observing"—given that Dennett admits we do come to be aware of our perceptions?[82] Naturally, we think it is the *self* that is observing our inner life. Dennett responds by denying that there is an observer of our mental content.[83] Asking "who" is doing the observing begs the question of an observer, a self. He responds: "All [perceptual] drafts are *products of*, not *candidates for*, interpretation" [my italics]. We are not separate viewers viewing

[77] Block, 1992, p205, also Aronson et al., Baars and Fehling, *ibid*., Teghtsoonian, *ibid*., p224, Van Gulick, *ibid*., p228

[78] cf. Velleman, p123
[79] Dennett, 1992, p195, p199
[80] Lycan, 1992, p216
[81] Aronson et al. 1992, p203
[82] Dennett, 1992, p221
[83] Dennett, 1992, p235

our experiencings, we *are* our experiencings.[84] I believe this answer is appropriate and deals with the second criticism.

Thus far, Dennett's argument does not seem to be obviously problematic for a Constructivist view with respect to the self (except for a question about distribution of responsibility, noted above). What his argument does, at least, is make a Cartesian self implausible. The next question is whether Dennett can dismiss the Cartesian controlling-self.

How Dennett tries to dismiss the Central Cartesian Controlling Self

We start with the assumption that Dennett is correct in dismissing the Central Observer. This does not contradict Constructivism. However, whether he can make the move from the denial of centralised apperception to the denial of centralised choice, is another important question.[85] For Constructivists do not believe that a componentised self bodes ill for moral responsibility.

At the end of his (1993) Chapter 8—on word choice, Dennett rejects the idea of a central word-chooser as well as a central perceiver. His argument runs (briefly) as follows.

A self, argues Dennett, (1993, Ch. 8), given the evidence of speech deficit disorders, is *not* an ultimate judge over choices; what happens is something more like a case of competing "candidates". He gives the examples of various forms of aphasia (*ibid.*, p249 et seq.) in which the person with the disorder struggles to choose the right word. The parallel he is drawing is with his "multiple drafts" of experiencing. Just as a person has many versions of what they experience, so they have many versions of what it is that they are going to say. What Dennett argues is that the symptoms of aphasia, of expressing an apparently random selection of words, is not the problem itself, rather, it is an indication of the normal underlying word selection process. In aphasia, what is happening differently than in normal people, is the person is simply not censoring or silencing their inner processing—*anasognosia* (p250). Thus, he argues, the processing of the words we "mean" to say is not as directed as we thought:

> "the brain's machinery is quite able to construct apparent speech acts in the absence of any coherent direction from on high" (p250).
>
> "'Just keep talking—Mumbling is fine'. Eventually the mumbling takes on shapes that meet with the approval of the author" (p245).

[84] as Hume said. See the later section on Hume's characterisation of the self.

[85] Let us not be tempted to say that denying physiological centralisation does not rule out the possibility of centralised mentality. The fact is that unless we wish to be dualists, each mental event must map onto some physiological event(s), and these mental events must exist in space and time as well—presumably at the site of these events.

Dennett then proceeds to argue that there is no reason to suppose that there is a particular directional function in the brain which coordinates the efforts of the subsystems, such as word choice. He says that although we are sometimes aware of what we're reasoning about, it is a rare experience, and generally we operate under a kind of automated process. Compare, for example, lace-tying, which is automated, and playing chess, which isn't. Once we know how to tie our laces, we don't have to think about it, and the same applies to most of our 'deliberate' acts—they don't involve any *deliberation*. This, Dennett claims, applies to all kinds of acts, not just speech acts. Rather, it is the deliberated kind of act which is the rarity (Dennett, 1993, pp251-2).[86]

If there is no central self which perceives, as he argues in earlier chapters, he continues, perhaps there is no central self choosing our actions and words either (1993, pp250-2). Further evidence for this emerges in the discussion of Dennett et al.'s (1992) article.[87] Jeannerod, for example, cites evidence that one avoids road obstacles before being strictly conscious of them, and that one only feels fear *after* the problem is avoided. Jeannerod provides further evidence with the same implication: An experimental subject is asked to simultaneously take an object and issue an utterance when he is aware of it. The utterance always comes after the taking (p212), which seems to imply that consciousness comes after movement was initiated (p213). Dennett eagerly responds: "Jeannerod ... demonstrate(s) disunity of the self" (*sic*) (Dennett, 1992, p235). It is questionable that this is what Jeannerod's evidence shows.[88] I think what it actually shows is that the consciousness of the events occurs *after the fact*. When Dennett declares, triumphantly, that this demonstrates "disunity" of the "self", I believe he may be exaggerating. What Jeannerod is highlighting is that the awareness of the brain states are just a separate group of states to *that* set of states that were controlling the body at the time of the motion decision. Actions, in other words, are non-conscious.[89] This is significant not so much because it shows disunity, but because we think of the self as choosing consciously.

The main problem with Dennett's discussion is that it is not clear, from his evidence (errors of speech)—that the existence of multiple mental states with similar content—entails the non-existence of some choosing "self" or "boss". To be explicit, Dennett's argument against centralised self-control goes something like this:

> a) Experimental evidence of timing delays between stimuli and awareness thereof indicates that information is processed as it becomes available, and that there is no

[86] See also Dennett, 1984, p87: We have to sometimes "relinquish control" in order to save time.

[87] Jeannerod, 1992, p212

[88] Dennett often intermingles two different theories: That the self has multiple perceptual drafts which involve processing delays, and that consciousness is epiphenomenal on brain states. This is an example of this kind of mix-up.

[89] Again, this confirms Libet's results.

single canonical version of any experience which could be the version that a Self could apperceive (in a central Cartesian Theatre). There are multiple drafts of each experienc*ing* and behaviour is modified as new information comes in.

b) Therefore there is no central perceptual arena, no Cartesian Theatre.

c) Therefore there is no audience in a Theatre, no central apperceiver or Self.

d) Experimental evidence of speech errors indicate that multiple drafts of choices of speech acts exist prior to enunciation or performance.

e) Therefore there are multiple drafts of any action choice prior to action.

f) Therefore just as there is no central apperceiver, there is no central place, (let's call it the Cartesian Courtroom) where candidates for action are "previewed" and decided between. Yet we normally think of our reasoning about decisions in this manner.

g) Therefore there is no central adjudicator, no judge in the Cartesian Courtroom.

I find the evidence Dennett provides to be compelling, but his reasoning questionable. He has a number of interesting anecdotes that suggest he is right. But much work needs to be done to *establish* it. I will omit this labour as I cannot think of a better defence of Dennett's ideas at this point. My conclusion therefore needs to be weaker: Dennett gives us reason to disbelieve in a centralised or singular controlling agent (self) in the mind, *and this violates our concept of the nature or role of a self, which is the entity that apperceives and makes choices.*

To summarise, Cartesian models of the self argue that there is a singular, unified, irreducible self in the mind which is responsible for making our decisions, and, that in order to *be* responsible for what we do, we would need such a self. Cartesian models however have two main problems, namely (a) whether there is such a self, and (b) whether such a self is needed for moral responsibility. Agreeing with the constructivists, to whom we will soon turn our attention, I believe that there is no such entity. But against the constructivists, I argue that such a self is indeed required for moral responsibility.

Section 4—Constructivist Models of the Self

Constructivists view the "self" as a term refering to the systematic relationships between the activities of the mind. Constructivists argue that the self is not a singular unified irreducible entity, but that it nonetheless is capable of causing our actions and taking responsibility for them. In his *Essay,* Locke provides an argument to the effect that that a person is individuated (identified) by his psychological continuity—a continuum of psychological properties,

memories and other mental states.[90] The self has a network of interconnected desires, reasons, and other mental states. Systematic relationships between these mental states manifest as regular behaviour patterns. That relationship of mental entities or states would be what a Constructivist would call the "self". When the collective effects of mental activities cumulate in action in some appropriate fashion, that is said to be the "choice" of the self. For Constructivists, talk of the self does not entail talk of a separate entity, or an irreducible entity. An agent, under the Constructivist view, is *in his actions* and can be responsible for them when his actions display the relevant systematicity, and an appropriate dependency on his mental states.

As we have seen, in order for a human action to be considered more than a mere movement, it has to be willed by a self. If it were not our *selves* which were the cause of our actions, they would, in a very real sense, not be *our* actions. There is a difference between our arms being caused to rise—a movement—and our choosing to raise our arms—in an action. If our actions were explained without reference to the role of our selves, our actions would lack purpose, be stripped of any dignity and be mere motions, as futile and accidental as a rock rolling down a hill (Jacquette, 1994, pp145-6). We would not be involved with the reasons for our actions, and the "act" would degrade into being a mere motion (Velleman, p125). An agent must *add* something to the motions (*ibid.,* p127). What makes us *agents*, says Velleman, is that we can "interpose ourselves" into a causal sequence, so that the resulting events *are* traceable to us (*ibid.,* p128). Velleman dislikes the way that many philosophers characterise our choosing process. The normal picture of motivation, says Velleman, runs like this. There is something an agent wants. The agent believes that a certain action will obtain that object of desire. That belief and that desire jointly provide the "reason" or justification for taking that action. This leads the agent to have an intention to perform that action, which in turn causes the agent's body to move. Velleman objects to this kind of description. This story, he claims, does not involve a *person* doing something, but rather is a discussion of a deterministic sequence of events. The person is just an "arena" where these events occur, his self plays no role (Velleman, p123). The person is just a "victim" of a series of causes. For a motion to be a bodily *action*, an *agent* must be involved as a person (see also Davidson, 1963, p700, for a classical discussion of the issue).

For Frankfurt, Watson, and Velleman, an agent makes a free choice if he or she, *qua* "self", is involved in the action through some specific feature of the mind, such as Velleman's *motive for practical reason* (pp142-3) or Frankfurt and Watson's respective *higher-order desires* and *valuational systems*.

Constructivism, using reference to these mental features, provides a model of what it is for an agent to act without requiring that the agent be some kind of separate entity that performs the action. We take everything in nature to be explicable in terms of cause and effect, without need to refer to special entities such as "agents" (Velleman, p129). The question, then, is whether choice is reducible to causal event descriptions (*ibid.,* p130). By what mechanism could a person "participate" in her bodily motions? (*ibid.,* pp132-3). Velleman suggests that this might be

[90] Hume's characterisation of the self is similar to Locke's.

achieved if the agent has mental states which are "functionally equivalent to a self" (p137). This suggests to me that Velleman wishes to isolate some mental system as *the thing in charge*, as providing the *functional equivalent* of a Cartesian self. The role of a "self", he says, is to "adjudicate conflicts of motives". The agent, reckons Velleman, is able to detach herself from her motives and bolster whichsoever she chooses to (p139). The agent is thus responsible for what she does, because through her own weight, as a self (*ibid.,* pp142-3), the agent determines the final outcome of the decision process, sometimes making the weaker motive *win*. The idea for Velleman, then, is that an agent is free just in case she uses the "weight" of her self—and "throws it behind" the choice which seems to be most reasonable, through the 'motive for practical reason'.[91]

Now let us compare Frankfurt and Watson's views. For Frankfurt, recall, the act is voluntary and free if it accords with the higher-order desires of the agent. So: we would say, simply, that if the agent acts in a way which accords with his higher-order desires, we would be able to say that the *self* of the agent was in his actions. The same applies to Watson. His model was that the agent would be free if his action were in accord with his values. So for Watson, the self of an agent is in his action when his action is in accord with his values. This is clearly the way in which we speak of peoples' choices.

So, if the motive for action is constructed from higher order desires, values, or motives for practical reason, then we say, of the body doing the motions, that the agent or self of that body is *in* those motions, and therefore the motions are freely chosen actions. Action reflects a decision of the agent's self when there is the right kind of relationship between the agent's subsystems (eg., the agent's mental states are devoid of unusual interferences), and when the actions display the appropriate kind of regularity, suitably described. In these circumstances (where we are said to act, or our self is said to be "in" our actions), our actions accord with our higher-order desires, valuational systems, and motives for practical reason. Selves are taken to be acting when actions accord with higher-order desires, valuational systems, or systems of reasoning. This is what marks an action as voluntary—originating in the self.

As long as we can characterise *when* a self is in an action or not, or reflects some desire, value or reason of the agent's, we can decide whether the agent was responsible or not *because her action reflected her desire, value or reason*. The "self" is therefore a term referring to the system of desires, reasons, values that an agent *has*, which leads her to do what she does. As long as her actions reflect her motives, her self is in her actions and the self is responsible (Leon, 2002b).

According to the constructivist model, a decision is made when the elements—such as beliefs and desires—are related to each other in the appropriate systematic way. The self, however, is not some separate fact over and above its components; when it has a relevant, appropriate,

[91] It is worth noting that this seems to contradict Hume's claim (which appears later), that the self cannot be separated from its mental states.

efficacious component, then it is said that the self is in the action or choice that has been made. This model of the self is plausible. It can accurately map onto our normal speech about the self, and when it is we say a self is in an action. But it has some difficulties.

Problems for Constructivism

I am concerned that constructivists are not providing an appropriate target for the reactive attitudes. Constructivism describes the self as the systematic relationship between mental states—but I am concerned as to whether we can truly say the *system* is causally efficacious. The causal powers inhere in the components; such as needs and desires, not, as far as I can see, in some over-arching system. This is not to say that constructivist models lack an characterisation of what it is for a self to be causally efficacious. We have seen this characterisation earlier on. I do, however, have a concern about *what* is really efficacious in our decision-making. Further on, when we get to the section on Parfit, this concern will become clearer.

Another potential difficulty with constructivism is that the kind of system it is proposing *may lack empirical support*. There may be better explanations for what we do, which may threaten the concept we have of a self as an integrated system. In other words, we may be more fragmented, more chaotic, and less systematic, than constructivists might be implying (cf. Dennett Ch. 8, 1993). For example, if Dennett was right, and causal states arose in competition with each other rather than in collaboration with each other, it would truly only be individual states which had causal powers, and there would be little room for any systematic relationships which indicated the presence of a self.

Next, consider the relevance of epiphenomenalism. Whether we accept Constructivism or Cartesianism, we do think of the self as being causally efficacious *qua* mental. If the self is a mental entity or system, and if epiphenomenalism is true, the self would be epiphenomenal. If this were the case, then *regardless* of whether we can talk of a self being in the actions or not, the self would not be efficacious *qua* mental. And a mentalistic self would lack causal efficacy if Libet's (1985) evidence was true.

Finally, allow me to express a logical concern that I have. Suppose we say a self is in an action when the action reflects the higher-order desires, or values, or reasons that the person has for acting. But if the term 'person' means a self, I believe it is circular to argue that to have a 'self' means to act in 'conformance to the higher-order desires, values or reasons that the self has'. If the self, on the other hand, merely consists in the evidencing of states, then the self is those states and cannot have them. To explain more clearly: to have a self means, on the Constructivist formulation, that (a) we have some states, and our actions evidence those states, and (b) our actions evidence that we have a self, therefore the self is those states. If this is correct, I believe that we can therefore have one of two interpretations of the Constructivist formulation of what it

is to have a self: *either* the self is evidenced by the regular actions of an agent and therefore the self is those states, *or*, if the self *has* those states and it is *irrelevant* whether the self is evidenced or not, the important part of selfhood would then be the mere *having* of those states. If the former interpretation were true, the self would not have any states. This sounds false. But on the other hand, if the latter interpretation were true—that selves *have* states and *that* is what it is to be the self, that would indicate that the self exists independently of those higher order desires, values or reasons. If the self exists independently of the higher order desires, values and reasons, then the self must be an entity over and above them, capable as it were of choosing which higher order desires, values, or reasons, to act on. This sounds like a more favourable interpretation, but if this is correct, then I believe this is a form of Cartesianism, which is not what I think a Constructivist would want.[92]

Section 5—The Skeptical Model of the Self

Skepticism is the view that there is no such thing as a singular, unified and irreducible self, and yet such a self is required in order for us to be morally responsible. If we lacked such a self, we would not be morally responsible for what we do.

> "If selves turn out to be [mere] constructs of information-processing systems, it will be hard to say why they are of any value ... I'm just something my brain made up. ... If people are valuable, it is not because they are imperishable souls connected to bodies for a brief sojourn. They have to be valuable for some other reason." (McDermott, 1992, pp217-8). [My insertion].

Dennett and Kinsbourne

The skeptical model derives from the evidence presented in Dennett (1993) and Dennett and Kinsbourne (1992).[93] In both those works, Dennett et al. present some skeptical concerns about our current conceptions of consciousness. The material presented uses some anecdotal experimental evidence which leads Dennett et al. to hypothesise that our consciousness is radically fragmented, lacking central apperception *and* central control. This work, while being based on somewhat anecdotal evidence, is nonetheless compelling. We have already seen Dennett's attack on the Cartesian model. What is interesting in Dennett, however, is his characterisation of the choosing process, which differs from the constructivist model. As we saw, he denied that there was a self which chooses particular words; rather, the brain puts forward various candidates, and the candidate which is strongest, by default, causes the final action. This

[92] Compare to Parfit's relay race, 1984, p223.

[93] Dennett himself is a constructivist, since he believes the self is not irreducible and he believes we are nonetheless morally responsible for our actions. I believe that Dennett's evidence points to the skeptic's conclusion; viz., that a reducible constructed self of the particular type Dennett describes *entails* that we are not morally responsible.

is not altogether different from what a constructivist model would hypothesise; but under that model, there would still be *a* self which chose between the candidates. On Dennett's model, there is no such process.

Recall earlier that we expressed a concern about whether a structured collection of states could be causally efficacious. For this debate, we first consider the work of Hume and Parfit.

Hume

Hume (*Treatise*, Book I, section VI) described the Self as follows:

> "For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception. When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of myself, and may truly be said not to exist."[94]

Hume is claiming that there is no observer independent of observations, within the mind; there is no "self" over and above the mental states in the mind. There is nothing that "has" our experiences as such; these experiencings just exist in our minds or brains—and we use this idea of "the self" to explain what "has" these perceptions.[95] Memory of our experiences, argues Hume, is the chief source of our sense of personal identity or selfhood. This sense of identity arises from our perceptions being interconnected by resemblance; or the "smooth transition" of "one idea into another". So we are like a commonwealth (Hume, pp299-302, p309, Parfit, 1984, p211).[96]

What is interesting in Hume's view is his view that the self is nothing over and above our sensory or memory states.[97] Effectively, Hume is denying that the self—or at least, a Cartesian self—really exists—eg., when we are in a deep sleep. As it stands, Hume's model might represent a problem for the idea that the self controls our bodies. Series of memories and sensory experiences don't strike me as the right candidates for being causal initiators in the motor cortex. We saw Searle raise this concern earlier (p501). Hume's model of the self and its

[94] *Treatise of Human Nature* BOOK I. "Of the Understanding", PART I. "Of Ideas, Their Origin, Composition, Connexion, Abstraction, Etc". Sect. VI. "Of Personal Identity." Also in Dennett, 1993, p412.

[95] Compare "it" in, "it is raining".

[96] Many writers responded to this to ask: Who is this "I" that is looking within and seeing only perceptions and thoughts—if not a self? (Dennett, 1993, p413). This is very like the Bishop Butler problem (cf. Parfit, 1984, p219). Butler argued that having mental states entails something that has them, thus mental states cannot be the thing that has them. It would be circular to make such a claim.

[97] This is what is meant by Searle when he refers to Hume's model as being a mere "bundle of experiences."

attributes would also have to include beliefs and desires if the self were to be causally efficacious. But since Hume does believe we are capable of choosing what to do, Hume must believe that we have beliefs and desires. So, since beliefs and desires are suitable as causal initiators, the self could be causally efficacious even if it is a Humean Bundle.[98]

Parfit

Parfit argues that identity (selfhood) is not a one-to-one relationship.[99] We could be something very different, eg., a whole team of "selves", coming and going out of existence, passing the baton of the mental contents from one mental entity to the next in a "relay race" (Parfit, 1984, p223). A person is somewhat like a club, with members coming and going all the time (Dennett, 1993, p423, cf. Parfit, 1984, p213). And Parfit argues that like clubs, "selves" are *empty* notions, they do not take objective answers about their continued existence. Suppose, Parfit says, that a club goes out of existence and is reconstituted later by some of the original members plus some new members. Whether that is really the same club is quite an arbitrary thing to decide; there is no *true* or *false* answer to the question of whether it is the same club. We could call it the same club, or just say it is a new club which is similar to the old. Parfit is arguing that the question of the existence of a club in terms of identity (ie., sameness) is indeterminate. Parfit is skeptical as to whether questions of identity comprise further facts (p242). This is not to say that Parfit does not think that selves exist. He feels that there do not need to be any absolute (yes or no) answers to questions like "is this me?" (1971, p3, 1984, p214). For example, he says, we do not ask, "is this the same nation?", when some citizen dies. He maintains that our attitude to ourselves should be the same: as long as some resemblance of our organisation is preserved in some recognisable matter by some appropriate causal means, then we have survived.[100] Personal identity, or the self, is similar to a club or a nation, in this respect (pp213-4). Rather, what matters, Parfit says, is if there is some kind of survival or continuity relationship (Parfit, 1984, p215, p256). Questions around identity are empty questions, because they only take an arbitrary answer. The answer we give is not given because it is right, but is given just because we feel obliged to give a determinate answer. Some cases will not admit of such an answer, and that is because there just is no such answer (1984, p255, 260).[101]

---

[98] but *only if* any mental states are causally efficacious. What interests me is not whether Hume's model can give us causal efficacy, but that Hume characterises the self as nothing over and above its states. This compares well with Parfit.

[99] One self to one body.

[100] There are real answers when it comes to questions of psychological continuity, but it is not clear whether there are real answers in all cases when it comes to questions of what we call "identity". This concern is precisely what I am alluding to in the matter of the self and its identity or sameness.

[101] For more detail on Parfit's argument and its supporting premises, see the relevant work.

This is the inference I wish to draw from Parfit: if we cannot give definite answers about the continued existence of a self or a club, we cannot give definite answers about what it has done or what it is responsible for. I believe that Parfit is right—that the self is *not* a "further fact", but I also believe that we *need* the self to *be* a further fact in order for us to give *definite* answers about its moral responsibility. What I am suggesting is that if the self is anything at all, a self *is* like a club, and parts of it do come and go, but insofar as we cannot get definite answers about the continued *existence* of any entity whose components change all the time, so we cannot get definite answers about what such an entity has *done*. This is not to deny the existence of a self, just to deny that some questions about it can be answered. I recognise, of course, that Parfit does not want to avoid the concept of moral responsibility, or have his model used to indicate that agents may lack responsibility. Parfit argues that the closer a "survivor" self is to the original self which performed an action, the more that survivor self can be held responsible for the actions of the original self. This is a reasonable alternative conclusion to draw, but it requires some additional things for it to work, eg., that the self would have to comprise a set of consistent states; beliefs and desires. Recall earlier that we argued that a self either *is* its states or it is a further fact over and above the states that it has. If a self *is* its states, then we cannot have definite answers about what that self has done, because its constituents change all the time. I do not believe this would provide enough stable consistent states, barring perhaps memories, to have anything sufficiently stable and consistent to play a suitable role as *the* one true causal effector. If however a self were a further fact over and above what states it has, then we could get a robust moral responsibility, for the self would not be in constant flux. But in that case, we would have a Cartesian self. And this is why Searle and I both believe that such a self is needed for moral responsibility.

The Skeptical Model

Having presented the evidence, we now put forward the skeptical model.

a) Only individual non-conscious brain states have causal powers and can lead to actions. These brain states might be informed by other states (particularly memories and learning), but ultimately the final action is more like a competition between various alternative brain states and their associated actions, and that the brain state which has the strongest influence on the motor cortex is the one which leads to action. There is no "deciding" between which states are to play out in action, rather, the states 'present' themselves to consciousness (or rather, reach Libet's *neuronal adequacy*), and whichever state is strongest by definition will be the one which causes the action. According to Dennett et al., the mind is just a collection of warring factions and separate drafts of decisions and experiences. This means that individual factions, drafts, mental states, *homunculi*, or brain units, whatever they may be, are the ones responsible for the eventuation of some or other choice. It is not a collaborative effort, no decision "comes together" at any point in time. A whole bunch of candidates offer themselves up for "selection" but ultimately one of them just wins out. Think of it as a relay race (Parfit, 1984, p223, Dennett,

1993, p238) rather than a tug of war. Certainly information is passed around and certainly there is internal interaction. But it is not clear that mental processing and decisionmaking is linear, coherent, coordinated, or particularly systematic. Rather, different candidates run towards a finish line, and ultimately one of them, the strongest, just does as a matter of being the strongest, get to the finish line first, which in our analogy, is the final action.[102] Likewise, it is not clear that there is any actual single entity or subsystem which at any point in time is the system controlling the whole. There is no actual "captain of the crew", there is no "internal Boss". Thus I suggest that the model presented here may be a more empirically accurate way to explain how it is that we act (Dennett, 1993, pp228-9).

b) Only single causal nexuses can be responsible for a particular event. I acknowledge that perhaps I have not provided a substantive reason why a committee-like, or structured self, could not be causally efficacious, and could not be the appropriate target of reactive attitudes. Why, in particular, would a functioning constructivist self not be such an appropriate candidate for selfhood? My intuition is that such a self does not have any overall *orchestration*, and that if anything, the 'self' merely refers to the interactions and relationships between states, and such interactions neither strike me as causally apt or appropriate for moral assessment. We think that what is appropriate for moral assessment is a person in charge of his feelings, not the arrangement of those feelings. It is not the car's engine components that are responsible for where the car goes, *per se*, we want to say the *driver* is responsible. And my argument is that there *is no driver*. Like with the question of Ultimate Responsibility (Kane), we need to settle on an answer as to ultimate causation.[103]

---

[102] This view derives from Dennett's Chapter 8, 1993, and his broad "multiple drafts" model, and the view that the brain states are non-conscious derives from Libet's (1985) evidence.

[103] Commissurotomy operations (the severing of the "commissures"—neural links between the two halves of the brain) are sometimes carried out to prevent epileptic seizures. If someone undergoes a split-brain operation, fragmenting his consciousness, there would be two separate consciousnesses, and consequently, arguably, two separate selves. But this may be evidence that we actually are (*pre*operationally) more than one person. People emerge from these operations and each side of their bodies behaves as if it were controlled by a separate person.

Marks (1981) maintains that if this interpretation is correct, the implication is that either we are always two people (and just don't know it), or, a new person is created in commissurotomy operations. Both alternatives are hard to accept. The fact that each surgical subject in a split-brain case evidences different experiences for each hemisphere, and the fact that neither hemisphere knows what experiences the other hemisphere *had*, indicates that we could have split selves. "Since there can be one state of awareness of several experiences, we need not explain this unity by ascribing these experiences to the same person, or subject of experiences" (Parfit, 1984, p251). This leads us to suggest that it is inconclusive as to which self is responsible for what the body does.

Marks objects to this. Splitting a person's brain like this does not entail lack of psychological unity prior to the split, or even *after* it. He claims that the singularity of selfhood persists *despite* experimental evidence that it does not. It is not clear that a split in mind entails a split in consciousness (Marks, p32). His argument, in brief, is that it is only through carefully selected experiments that we can detect disruptions in personhood of split-brain cases (p41). He seems to think that as long as both halves of the brain (a) retain some communication—which they do through the lower brain, and (b) as long as they both retain sufficiently similar characteristics, behaviours, goals, etc., that the self is not really split. If a person's behaviour postoperatively is largely consistent with his preoperative behaviour, Marks argues, then we are obliged to explain it by a largely unified mentality.

We therefore need a single unified irreducible self if that self is to be responsible for the event that has been caused. Otherwise the responsiblity belongs to the individual mental state(s) which played the causal role(s). We do tend to want to blame *one* person, and there is evidence that indicates that the concept of "one" person is open to some doubt. This requirement of a single causal nexus is an intuitive committment I share with Searle. This does not mean that the intuition is correct, but it is clear that we at least believe that the self as an entity or system is a thing that is responsible for what we do, and that leads me to believe that we do intuitively want to attribute causal powers to a single entity or system acting as a *whole*. My skepticism is thus around not only whether the self really exists, but also around whether it is the sort of thing that could be a causal nexus.

If we want to attribute causal responsibility to a self, we have to defend the view that the self, in virtue of being a complex structure, can *as a whole* be a cause of an action. This concern originates in Ryle (1949, p17)—the "Category Mistake". When "the University" makes a decision, it usually means that the individuals on Senate have voted for a majority in favour of that decision. Suppose the University's senate takes a decision, and suppose one of those members of senate who voted in favour of that decision leaves the University just as the decision is taken. Suppose now that the person who left was also the tie-breaker; his vote caused the senate to have a majority and as a result, the decision was taken. Does that mean that the senate did not take that decision? No, we want to say. The senate still took the decision despite the fact that the person who caused an end to the tie, left. But at the same time it is clear that the individual who left was needed to make that decision. The question then is this: do we attribute senate with responsibility for the decision even though the element that caused the decision has left? Many people, I think, intuitively, will still argue that yes, senate is still responsible. I however have the opposite intuition: namely that the individual who pushed the vote forwards was the one who was responsible, and that ultimately it was his decision. The 'self' is like this.

There may be a fallacy of composition here: Just because some mental state can cause an action, it doesn't mean the whole superstructure of the self is causally efficacious. We would therefore only be able to rescue the role of a "Self" here if we could show that the self *as a whole* was token-identical with that brain event.

c) The 'self' can have two plausible meanings or interpretations. Constructivism says that we have a systematic self, and we have behavioural regularity; thus the self is evident in the actions. I prefer to say the opposite; not that we have a self and it causes our behavioural regularity, but rather, that we infer that there is a self because we see behavioural regularity. In other words, apparent behavioural regularity does not entail that there exists some regulatory self, or some regulated system. If we do have some behavioural regularity, it may indicate the presence of a regular cause, but that cause need not have the properties we traditionally assign to selves, such as consciousness, moral responsibility, freedom of will, etc.

d) Since we lack a single regular standard causal nexus in the mind—a self with causal powers—it is not the self that is causally responsible for our actions, but the individual brain states. Thus, since moral responsibility requires a self that controls the mental states and chooses which shall be efficacious, and since we lack such a self, our selves are not responsible for our actions. This is the logically necessary conclusion of (a) through (c). Yet we require a self which is responsible for our choices if we ourselves are to be morally responsible for our choices. Therefore we are not morally responsible for our choices. Therefore we lack free-will.

Conclusion

The reader might want to agree that "obviously" there is no central controller, and wonder why I should feel that it is necessary to argue against it. The point is *that is how people see themselves*. Certainly I did, until I read Dennett. The "self" is traditionally construed as a centralised single entity, the personhood of a person, it has our perceptions, memories, and makes our choices. But the endeavour of this chapter has been to show that the evidence is *against* such a thing.

> "... the strangest and most wonderful constructions ... are [those] made by ... *Homo sapiens*. Each normal individual of this species makes a *self*. Out of its brain it spins a web of words and deeds, and, like the other creatures, it doesn't have to know what it is doing; it just does it... So wonderful is the organisation of a termite colony that it seemed to some observers that each termite colony had to have a soul. ... We now understand that its organisation is simply the result of a million semi-independent little agents, each itself an automaton, doing its thing. So wonderful is the organisation of a human self that to many observers it has seemed that each human being had a soul too: a benevolent Dictator..." (Dennett, 1993, p416).

I am not suggesting that Constructivist models are wrong. Our selves are indeed systems constructed from mental states, and it is meaningful to say that a "self" is "in an action" when the action conforms to some regularity of values, reason, etc. What worries me about this is that it seems to fail to answer the question. The question is this: Is there *an* entity which alone is causally responsible for the actions of the body? The answer is *No*. The only way to have moral responsibility is to have a particular "further fact"—an entity in charge of what we do. But if the self just is its states, then there *is* no overall "control mechanism" which over-arches and determines the interaction between mental states.

> "If a self isn't a real thing, what happens to moral responsibility? One of the most important roles of a self ... is as the place where the buck stops... If selves aren't real—aren't *really* real, won't the buck just get passed on and on, around, forever? If there is no Oval Office in the brain, ... we seem to be threatened with a ... bureacracy of *homunculi*, who always reply, when challenged: 'Don't blame me, I just work here'. " (Dennett, 1993, pp429-430).

CHAPTER 6
Whether Free-will Matters

Many writers believe that the idea of free-will is important, and without it, we stand to lose a wide variety of desiderata which are crucial to our lives, cultures, and way of interacting. This chapter discusses this matter. In the first section, we consider whether or not having free-will obtains any of these desiderata, and therefore whether having free-will matters. In the second section we speculate about why people believe in free-will anyway.

Section 1—Whether we need free-will and whether it matters

It is apparent that if we do not have free-will, certain important features of our moral lives, or desiderata, would fall away. We generally believe that we need free-will to justify *the system of morality, the reactive attitudes*[104], *law and accountability*[105], *personal efficacy*[106], *actions based on reasons*[107], *distinction from the animals, human dignity*[108], *creativity, due credit, autonomy, just deserts, individuality, life hopes,* and *love*[109]. Pereboom (2001, xii) says that it may be possible to meet some of these desiderata in the face of determinism. I will be defending this position in this chapter, viz., that despite lacking free-will, many of our desiderata will remain intact. I will also consider whether having free-will would matter, depending on the extent to which free-will is required to preserve some desideratum.

1.1. Morals

Of all the desiderata that free-will purportedly secures, none concerns people as much as morality. It is said that if we lacked free-will, we would lack morality, and therefore, having free-will matters. I am not sure this is the case.

In this section, I will make two arguments. The first argument is that even if we have free-will, it will at most preserve causal responsibility; it will not guarantee we are also morally responsible. The second argument I will present is that even if we lack free-will, we could still preserve some conception of moral conduct, even if we are not ultimately responsible for how we conduct ourselves.

---

[104] P. Strawson, p5 et seq., and Wallace, p9 et seq.
[105] Van Inwagen, 1983, p18-20, G. Strawson, p15, Dennett, 1984, p153
[106] Van Inwagen, 1983, p18-19, p134
[107] Leon, 1997, p5, Glover, p466, Wolf
[108] Frankfurt, p92 in Watson, *Free-will*, p80 et seq., p82
[109] Kane, p80

I do not believe that having free-will would necessarily secure moral responsibility. Consider this. For a moral claim to be true, says Double, it needs to have non-linguistic, non-attitudinal, non-conventional ontological grounding. Double asks: *are moral properties "part of the furniture of the world"?* (p153).[110] He discusses some models of what moral properties could be, and ultimately rejects them (Double, p151 et seq.). If it were true that there were no moral properties, then no event could have moral properties. Now, suppose this were true and yet we agree that persons had free-will. If this were the case, then all actions, despite being free, would not be moral. Thus, it may be possible to have free-will and yet lack moral responsibility. If moral responsibility is what matters to us, then having free-will would not matter as much as we thought, since having free-will would not guarantee moral responsibility.

Let me draw a distinction in the meaning of the term "responsible", which many writers fail to make, which may make it clearer how morality and free-will can come apart.

There are two different senses of "responsible for": *moral* responsibility and *causal* responsibility.

> An agent is *causally* responsible if some states in the agent, cause some movements of the agent's body, which lead to some event occurring, without which the event would not have occurred. An agent is causally responsible when the agent is the cause of some event (regardless of its metaphysical properties).
>
> An agent is *morally* responsible if that agent was causally responsible for an event that has moral properties—a "morally relevant" event, and that agent deserves reactive attitudes as responses to her actions.

In order to be *causally* responsible, there would have to be certain factors present. Agent *A* would be *causally* responsible for event *E* if the following criteria were met:

> a. *A* wanted *E* to occur *and*
> b. *A* acted to cause *E* to occur because of his/her own desire and because *A* had a properly formed belief as to how to realise *E and*
> c. *A*'s actions were sufficient for *E* to occur *and*
> d. *E* actually occurred (because of *A*'s desire), in the right kind of way

Propositions (a) to (d) represent a simple formulation of free-will. Thus, in itself, having free-will entails being causal efficacious. Now, in order to be *morally* responsible, not only would points (a) to (d) have to be true, but the following would *also* have to be true:

---

[110] The first thesis that Double postulates is that the term "free-will" is so incoherent, and so hard to characterise, that in accepting any particular characterisation, we rule out the possibility of accepting some other characterisation. Double argues that there is nothing that properly answers to the concept of free-will (Ch. 6). I will not discuss this view of his further, or the debate around moral realism, for reasons of space.

e. *E* is a morally significant event (it is appropriate to label it as "bad" or "good"), and *A* deserves consequences or reactive attitudes for causing *E* [111]; the action is credit- or blame-worthy.

My skeptical point, then, is this: Point (e) is required for agents to be *morally* responsible, but it is not obviously entailed by (a) to (d). If (a) through (d) is a plausible minimal account of free-will, then it is apparent that free-will does not necessarily secure moral responsibility or desert. (Double, p217, 224). Free-will might be *necessary* for moral responsibility, but it is not *sufficient*, and thus it does not completely secure something we value. Therefore, *even if we had free-will*, we require moral truths to secure moral responsibility, and thus, free-will alone would not be able to give us one of our desiderata.

Let's consider the second argument. Suppose now that we lack free-will. Obviously, if free-will is the only thing that justifies morality, and we lacked free-will, we would lack moral systems. But if this were the case, it would not be that free-will did not matter, it would just be that we lacked something that did matter. I do wonder, however, whether or not it is possible to reconstruct some form of moral system in the light of these considerations.

Pereboom tries to show that we might be able to rescue moral systems even in the absence of free-will. He argues (against Double), that even if we lack free-will, and even if we lack moral responsibility, that does not mean that we need abandon all forms of moral systems. Pereboom says that there are two aspects to moral dicta: *reactive attitudes* and *ought-statements*. He argues that if hard determinism were true, we may still be able to preserve ought-statements but not the reactive attitudes. The problem, he acknowledges, with this position, is that we generally take it that "ought" implies "can", and if hard determinism threatens "can", then ought-statements would not be legitimate, since there would be no point to holding people morally responsible if it was not up to them as to whether or not they did the morally right thing (p143). But even if we can't adhere to ought statements because of the lack of alternative possibilities, there could still be things that are truly good or bad (see eg., Pereboom, p147). More importantly, Pereboom argues, moral judgments can guide actions and act as causal determinants towards the final decision. Moral judgments can thus help in the decision-making process (p148). This argument is interesting, because it seems to indicate that moral edicts could still have psychological force on an agent, and thus, we may indeed be able to preserve some aspects of our system of morality.

Pereboom also argues that we could resort to virtue models of ethics; ie., that a person develops their character according to how she practices the ethical virtues (p150). There is nothing in hard determinism *per se* to preclude a person being able to be a certain way, and controlling him or her self through practice. Hard determinism also doesn't rule out the possibility of presenting

[111] Some writers argue that merely intending to be the cause of an event is sufficient to be responsible for it, however, we omit this debate here for reasons of space.

oneself with moral reasons for doing things, and examining one's past to see what kind of person one is, or is becoming.[112] Pereboom compares it to developing a work of art.[113] (p153). Thus neither hard determinism nor incompatibilism necessarily rules out all systems of ethical practice. I am sympathetic to this argument. It is reasonably clear that when one acts, one develops one's personality. However, this is not the same as "freely choosing", and nor does it make one *ultimately* responsible for who one is.[114] So, in agreement with Pereboom, we may be able to preserve some form of moral or ethical conduct, without necessarily being *also* morally responsible for that conduct. The questions is whether it is the moral responsibility (creditworthiness or blameworthiness) that we value, or whether it is just that we value that people behave morally. I suspect it is the latter that is more important.

In an earlier chapter, I have argued that if we lacked free-will, then we would lack the autonomy to be responsible for how we change our selves or control our selves, for change and control would come from prior deterministic circumstances.[115] But this does not mean we *cannot* control our-selves. It is clear, for example, that some brain states *can* prevent other brain states from having an effect (see Libet's "Veto"). Thus, differently understood, autonomy would in some sense still be possible. Hard determinism, *per se*, is therefore not incompatible with the possibility of self-control, since the control of the self would be carried out through a deterministic process, which need not be free. But recall now, in a previous chapter, I argued that there is no substantial notion of 'the self' which we can use to give persons moral responsibility. If we construe autonomy as self-control, and there is no self, and control is not consciously effected, the reader may be curious as to why I believe autonomy can be preserved. What I have done is provided a form of autonomy in which the mind, or brain, or body—whichever you prefer—is able to modify behaviour through deterministic systems (whatever systems they may be)—and that is a form of self-control or self-regulation. It doesn't require an overarching self of the sort that I attack in that previous chapter. Thus again, hard determinism does not necessarily rule out moral systems—especially systems of ethical conduct. At most it rules out blameworthiness or creditworthiness.

In this section, we have seen two arguments with two conclusions. Firstly, we have seen that free-will does not guarantee that we are morally responsible, because of the possible correctness of moral skepticism. There is thus reason to doubt whether having free-will would preserve moral responsibility. Thus even if we had free-will, it would not necessarily matter because it could fail to preserve something we value—moral responsibility. Secondly, we have seen that

---

[112] I recognise that Libet's evidence may complicate the picture here, so I ask the reader to withhold judgment at this stage.

[113] This is very reminiscent of Sartre and Kane's characterisations of self development through self-forming acts.

[114] See G. Strawson (1994), for the argument for this, or refer back to the chapter on Incompatibilism, where Kane gives an argument for the necessity of ultimate responsibility.

[115] This was discussed in the chapter on Incompatibilism under the heading "causa sui".

even if we don't have free-will, it may still be possible or reasonable to have systems of ethical conduct. If we only care that people are able to act morally, then it may be possible to retain some system of morals, even if hard determinism is true. Thus perhaps free-will does not matter as much as we thought it did, because we do not need it in order to preserve some moral systems.

Let us now consider whether free-will preserves the reactive attitudes, and if it does not, whether there would be any point to the reactive attitudes.[116]

1.2. Reactive attitudes

Introduction

It is believed by many current philosophers that the reactive attitudes, ie., those emotions and reactions we experience in regard to morally pertinent events, are justified or made legitimate by the fact that the person who is the target of these attitudes, is the *appropriate* target *in view of their having free-will*. Since being an appropriate target of the reactive attitudes is tantamount to having moral responsibility, we could find that having free-will would matter if it justified the reactive attitudes. I will argue in this section that this model is incorrect, ie., it is *not* that

> the person has free-will, and freely chose to do an action, and we therefore justly express reactive attitudes to that action,

but that:

> we *impute* free-will to the agent who is the recipient of reactive attitudes, *because* we are (naturally) experiencing those reactive attitudes; it is not free-will *per se* that matters in justifying the reactive attitudes. This will be argued in detail below.

Let us discuss these two opposing explanations for the reactive attitudes.

Discussion of the Reactive Attitudes

The conventional model

On the conventional model of the reactive attitudes, there are conditions under which the reactive attitudes are appropriate or inappropriate. The conditions under which the reactive attitudes are said to be 'appropriate' are just those in which an agent freely adheres to, or violates, the known

[116] See Wallace, p51, where he defines morally responsible as meaning that one deserves reactive attitudes.

and morally correct code of conduct. This code prescribes certain reactions for certain actions.[117] When we say someone "deserves" something, it means that in some sense she is an appropriate recipient of some reactive attitude or consequence (Fischer, 1986, in Double, p75, p136), because, through her own actions, she has made herself an appropriate recipient (P. F. Strawson, p7, Wallace, p78).

Let's take an example of the normal construal of the circumstances under which the reactive attitudes are justified. Suppose it is in fact wrong to harm someone, and suppose the agent in question does harm someone. And suppose the agent knows it is wrong, and that the agent knows she was harming someone. Suppose also that the action was under her control, sensitive to her reasons, and she could have done otherwise.[118] When these conditions obtain, it is generally thought that it is intelligible as to why the agent should receive a response of the reactive attitudes. Under these circumstances, the person who was wronged is said to be entitled to be angry, or to have a reactive attitude of resentment; that the action was wrong, and done deliberately (knowingly), makes the reactive attitude appropriate. But my question for those who think the reactive attitudes are appropriate under such circumstances as these, is: are the reactive attitudes, in such a case, actually appropriate? Let's discuss this.

In *Freedom and Resentment*, P. F. Strawson discusses the reactive attitudes, and suggests that the truth of determinism—ie., the free-will debate—is irrelevant to the reactive attitudes. Moral responsibility, Strawson suggests, is based on the extent to which we deserve reactive attitudes in response to our actions. This is a compatibilist position, because it does not assert that determinism renders free choice impossible (Pereboom, xvii). Strawson argues that what matters to us is how we are regarded and treated—whether with respect or contempt, goodwill or malevolence, etc. For example, if someone seems to want to harm us, we experience greater displeasure than if she harms us accidentally (P. F. Strawson, p5). The reactive attitudes thus involve interpersonal regard (p16). If someone is in some way morally underdeveloped, we adjust our attitudes to him and treat him differently (pp8-9). However, he continues, "to adopt this objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject... for treatment; as something certainly to be taken account ... of; to be managed or handled or cured or trained..." (p9). This way of treating people is the one we resort to when someone seems to be morally compromised or underdeveloped. Would determinism perhaps lead us to adopt this attitude to everyone? (pp10-11). We may think it would do just that, but it is "practically inconceivable" that we implement such an approach. Strawson suggests that our committment to ordinary relationships is so "deeply rooted" that we could not become fully objective in this manner about everyone (p11). Strawson argues that it is not the case that we use the objective attitude towards those who are not "morally developed" because we believe they in particular are determined—rather, we adopt that mode of operation because in

[117] Generally, we exclude the mentally disabled, animals, and young children from such desert, and this is because we assume they have diminished responsibility, or lesser freedom of will through having lesser knowledge or rationality (cf J. S. Mill).

[118] perhaps in the conditional sense, in which had her reasons been otherwise, she would have done otherwise.

some way, such people are *beyond* the normal reactive attitudes; it is because of how these persons are determined that we adopt different attitudes to them. But Strawson claims that we nonetheless cannot avoid having the reactive attitudes (p18). "This commitment is part of the general framework of human life, not something that can come up for review" (p13).

Merely describing punishment as a means of social regulation, as something objective to be handled objectively, strips from both the people involved—the victim and the perpetrator, the real emotional matters that are at play: the emotions of guilt and resentment (P. F. Strawson, p20). It would involve a deep conceptual change in our world (p21). "...Attitudes and feelings... form an essential part of moral life" (p23)—our lives would be impoverished were we to abandon them. But Strawson does also recognise that we do treat a person of diminished moral capacity differently, 'objectively'. I realise Strawson is merely trying to disentangle the explanation for reactive attitudes from the traditional free-will debate. It is clear that he thinks that there is a connection; he argues that those who are mature moral reasoners (eg., adults) would deserve reactive attitudes, and we can argue, with Leon and Wolf, that those who reason about what they ought to do are, in so reasoning, acting freely. Strawson feels that if a person's circumstances are such that they have no moral capacity, we justly treat them 'objectively'.

Wallace, whose ideas are a development of Strawson's, seems to hold a similar position on the question of determinism. "Determinism is not a genuine threat to our practice of holding people [morally] responsible" (Wallace, p8). If determinism were true, says Wallace, we'd neither need to, nor want to, drop the reactive attitudes (pp8-9). So Wallace's position is also compatibilist. We might consider only giving up *those* reactive attitudes, which, because of pragmatic considerations, make things worse (p9). The reactive attitudes are connected with expectations—of our selves, and others—expectations about conduct and its significance. The reactive attitudes are about not only interaction with others, but our expectations of others, and in particular their treatment of us (pp26, 31). Thus they are part of a known moral code, a code of conduct. "To hold a person morally responsible... is to hold a person to the moral expectations that one accepts" (p51). We need to ask what the point of moral reactive attitudes are, and how they go beyond the action that was done (p54). This connection of the reactive attitudes with morality, is a matter of their relation to behavioural obligations that we accept (p77). In other words, we react to others, and their failure to fulfil our moral expectations, if we (and the others) know of some moral code or moral expectations, and one of us fails to fulfil the expectations of that code. But this raises a question of why we should then bring in the emotions, if it's merely an objective matter of an objective violation. Wallace contends that unless we do so, it would leave out what is important in blame: the emotions; which agrees with Strawson (p20). Treating people 'objectively' would reduce blame to merely being a description of what was done and the implication that it was not acceptable. Blameworthiness, Wallace maintains, does require a belief in the *appropriateness* of a reactive emotion (p78). And if the person has been unable to use her power of rational self-control, she should be exempted from blame (Wallace, p157). Thus, both Strawson and Wallace seem to be saying that the reactive attitudes are some kind of instinctive behaviour, but that they accord with a known code of conduct. Their legitimacy or

appropriateness derives from the extent to which an agent has adhered to or violated that code freely, and the question of whether determinism is incompatible with free-will has no bearing on our deserving the reactive attitudes (Strawson, p5, Wallace, p8). Thus, for Strawson and Wallace, free-will matters because it distinguishes a morally responsible person, who deserves reactive attitudes, from one who is morally incapacitated.

Critical Discussion

a) Reactive attitudes and determinism

Both Strawson and Wallace think that there are certain conditions that need to obtain in order for the reactive attitudes to be appropriate, and vice versa. But neither writer thinks that determinism itself is a condition which makes the reactive attitudes appropriate or inappropriate, so much as the *kind* of determinism. That is where I depart from their view. Wallace and Strawson argue that determinism does not rule out the reactive attitudes (Strawson, p5, Wallace, p8). It seems as if Strawson and Wallace want to say that the debate around determinism is not relevant to our exhibiting the reactive attitudes. I do not think, however, that they are being entirely consistent on this, because they both recognise that morally challenged persons require special treatment, and this is because such persons are determined by circumstances which exempt them from moral responsibility. How someone is determined is at least relevant in ascertaining whether an agent ought to be reacted to. While particular determinants, such as mental retardation, may inhibit our free-will and exempt us from moral responsibility, Strawson and Wallace probably would not argue that it is because of determinism *per se*, but rather because of the particular type of determinant. But I believe I have argued in a previous chapter that if hard determinism is true, then all determinants are relevant, not just ones which pertain to an agent's capacity for moral judgment.

b) Whether free-will really matters in justifying the reactive attitudes

The reactive attitudes are said to be inappropriate under certain conditions, eg., where a person who did something did not have a free choice, or she lacked moral capacity or maturity. Under such circumstances, as Strawson observes, we treat the person objectively, and exempt her from reactive attitudes. An agent does not deserve reactive attitudes when that agent lacks free-will. Symbolised, where FW is "the agent has free-will", and RA is "the agent deserves reactive attitudes", this is: [~FW ⊃ ~RA].

Thus, since the reactive attitudes are inappropriate under some circumstances, we take it that the reactive attitudes *are* appropriate under *other* circumstances. What circumstances could those be? Well, just those circumstances under which a person who is appropriately mature, and capable of moral judgment, freely performs some act. Under those circumstances, the person

would be the appropriate recipient of the reactive attitudes. An agent does deserve reactive attitudes if that agent has free-will. Symbolised, this is: [FW ⊃ RA]. And it does seem as if [~FW ⊃ ~RA] entails [FW ⊃ RA].

If the capacity for free-will is one of the conditions that makes sense of the reactive attitudes, then having free-will matters, because without it, we could not justify the reactive attitudes. But I remain skeptical. The reason I still doubt this is as follows. In the converse argument, namely [~FW ⊃ ~RA] it is not only true that [~RA], it is also true that [~X, ~Y, ~Z], where X, Y and Z are other states of affairs which just so happen to be false in this possible world.

In other words, I wonder whether a free agent *ought* to be a recipient of the reactive attitudes. I suspect that free-will alone cannot justify the reactive attitudes. We need free-will *and* a truth that says that the reactive attitudes ought to follow a free action. The standard position is that if you are freely responsible for some act, that you are *therefore* eligible for some reactive attitude. The question I pose here is around this "therefore". *Should* we even get the reactive attitudes as a response to our actions? If a person performs an action under duress, and she was not free in performing that action, and we do not react, it doesn't entail that we ought to react under circumstances when she performs an action freely. It is not obvious that we ought to react when the person does something freely; all sorts of other possibilities are available. I don't see how we can get from a belief that someone has "deliberately" done something to our everyday acknowledged practices of reacting. I recognise we can read "appropriate" as "intelligible". My concern is with *justification*. Suppose that a reactive attitude is merely said to be "intelligible". If a reactive attitude is intelligible, people might take that as license to mean that it is also justified, or, that it ought to be expressed. Just because a reactive attitude might be intelligible in a certain context does not, however, entail that the reactive attitude is justified or *ought* to be expressed. Thus again, free-will does not entail the expression of reactive attitudes.

Furthermore, I have argued that if hard determinism is the case, we must lack free-will. Thus, since [~FW] is always true, [~RA] is always true. If that is the case, then we have lost something we value: the right to have reactive attitudes as a response to persons who freely choose to perform some act. If there is no free-will, it would entail that we have to always treat people 'objectively', as Strawson suggests.

c) If we lack free-will and lack justification for the reactive attitudes, what then? Reconstruction.

Both Strawson and Wallace seem to wish to legitimate the reactive attitudes, at least in part, by appealing to human nature (P. F. Strawson, p13, Wallace, p78). Hume apparently also defends this view (in Pereboom, p91). I agree. I believe that the reactive attitudes are merely *instinctive* reactions, originating in the *"fight or flight"* instinct (Sdorow, pp466-8). If this is correct, reactive attitudes are to be explained psychodeterministically or neurologically rather than as moral entitlements. But this does not provide the reactive attitudes with moral force—in other

words, just because we naturally experience reactive attitudes, it doesn't mean we *ought* to express them. This would be a case of the *naturalist fallacy*. I agree with Strawson that the reactive attitudes are not the sort of things we can push aside with "intellectualism". But, lacking a moral justification, the reactive attitudes can only be justified as *having a role as behaviour modifiers*—practices we employ to control or shape behaviour[119] (see Skinner, *Walden Two*, for a similar view). But if this is correct, we see that regardless of free-will, we could possibly retain some semblance of moral systems. To give an example; instead of saying someone is "doing something evil or wrong", we could reconstruct this as "doing something harmful—biologically destructive", and instead of "punishing" the person, we would "decondition" his behaviour, perhaps through recourse to reactive attitudes.

Harsh punishments are not appropriate if some form of hard determinism is true; rather, preventative detention or rehabilitative programmes should be followed (Pereboom, xix-xx, p96). We do not hold earthquakes morally responsible, so why should we hold people responsible if they too are merely deterministic in their motions? (p154). But Pereboom suggests that moral reasons can act as determinants and exert force in our decisionmaking. This is true. If hard determinism were true then moral reasons would reduce our free-will, because then moral reasons would be mere determinants. We might continue to treat people as if morally responsible and blameworthy in order to reform them, so they don't persist in immoral behaviour (p156). We could just make use of morals as leverage, eg., through admonishment or encouragement (p157), for the purpose of shaping society. My reconstruction here, then, is not focused on Strawsonian treating of persons as a subject, an agent with personhood, but rather my reconstruction aims to justify why we could retain the practice of expressing reactive attitudes without the moral baggage that usually accompanies them. Reactive attitudes should now be seen as instrumental toward a greater goal of the general prosperity of society.

Pereboom (pp200-2) discusses various reactive attitudes and concludes that only the harmful or aggressive attitudes, generally, would fall away because of hard determinism. Those reactive attitudes which do not presume the existence of free-will, for example, pity, admiration, etc., may survive.

Let us briefly follow Pereboom's attempts at reconstruction of the reactive attitudes in the context of the possible truth of his view. Pereboom considers praise and blame, and argues that these are in fact undermined (p140) if hard incompatibilism or hard determinism is correct, since choices are always determined by external sources, or are random. He tries to reconstruct praise in terms of gratitude. Gratitude can be reframed as a feeling of joy at something having gone one's way without one's intervention. The same applies to love; we do not need to believe that the person is responsible for choosing to love us, or meriting praise, or whatever. For Pereboom, love is merely a feeling of wishing well for the other person, joy at how they are and a desire to be with them. This is also not threatened by hard determinism. Pereboom then turns to the

[119] Wallace mentions but disputes this view, (p54), saying it leaves out the attitudinal aspect (p56), as does P. F. Strawson, p20.

question of blame and guilt. He cites an example from Wolf in which only those who do right are praised, and those who do wrong are not blamed. The reason is that while both are mere expressions of feelings, blaming typically causes pain, and thus cannot be right to do unless it is deserved, which is ruled out by hard determinism (p141). Pereboom then discusses forgiveness, and suggests that though forgiveness seems to entail true free choice of wrongdoing on someone's part, it can be reconceptualised as an arrangement in which the offending party recognises the problem and resolves to remedy or change it, and the offended party recognises that willingness and ceases to feel negative towards the offender. The reactive attitude of blame alone falls away, if hard determinism is true. Guilt and repentance, he counterposes, are however useful for self-modification and moral improvement, where we learn from having done something wrong and use these feelings to motivate ourselves to do better in future (Pereboom, p204). Yet these feelings may be thought to be undermined by hard determinism. But Pereboom argues that we do not need this. All we need is a recognition that what we did was wrong, and a desire to correct it. We do not need to torture ourselves on top of it. Perhaps we could seek therapy, instead of constantly reframing ourselves as bad persons. We accept that we did something wrong and learn from it, as Dennett says (1984, pp142-3).

We have argued that free-will might have nothing to do with justifying the reactive attitudes at all; they are justified, or are appropriate at most in virtue of being natural, and having instrumental use as behaviour modifiers and tools of social control. Therefore free-will does not matter if we wish to preserve the reactive attitudes.

1.3. Law and the penal system

Thus far, we have been discussing whether free-will would preserve our desiderata. We have seen that there is reason to believe that free-will alone may fail to preserve some of these desiderata, and we have tried to show that the desiderata could in some cases be preserved even if we lack free-will. If we had free-will, it would justify the legal system, and therefore, having free-will would at least matter in that regard. But given that we have argued in preceding chapters that we do lack free-will, it looks as if the legal system, which seems to require free-will, may be threatened. In in this section, therefore, I will argue that we can justify the legal system by means other than appealing to free-will, and because of that, I argue free-will does not matter—since we can use other theories to preserve what matters to us. Let us start by discussing how legal systems require free-will.

If we had free-will, it would be possible to hold us causally accountable for what we do; that is, if an explanation is required as to who caused some legally problematic event, it would be possible to find someone who was causally responsible for that event. Suppose, however, that the penal system *depended* on the truth of free-will and moral responsibility. And suppose these were disproven. If this were the case, then the penal systems would not be justified. We would not be entitled, morally or logically, to punish. Legal systems operate on the assumption that a criminal

does something that is actually wrong and they do so knowingly, intentionally, willingly, and freely. If criminals lacked free-will, or if morality was not real, then criminals would have to be excused.

Consider the American legal system as an example. That system has an "insanity defense". If someone commits a crime and can argue he was insane at the time, he can be acquitted. There is a "cognitive rule" (about knowing right from wrong), and a "volitional rule", which "presumes the reality of free-will"—meaning that a person is insane if he lacks self-control (Sdorow, p582). Assuming we have seen some reasons thus far to disbelieve in free-will, the American model of legal accountability might need revision, for under American law, criminals would always be able to make an appeal to the insanity defense (because they lack free-will). Consider now, the South African system. In order to prosecute a person under South African Law, a "delict" must be found:

> "To found liability, further requirements must be met. These requirements... appear from the following definition: *A delict is the act of a person which in a wrongful and culpable way causes harm to another*. All five requirements or elements, namely an *act*, *wrongfulness*, *fault*, *harm*, and *causation* must be present before the conduct complained of may be classified as a delict. If any one (or more) of these elements is missing, there is no question of a delict and consequently, no liability." (Neethling, Potgieter, Visser, 1996, p4) [Neethling et al.'s italics].

If we had free-will, we would indeed be able to preseve a legal system such as this. But if we lacked free-will (as we have argued in previous chapters), we could not obtain a *delict*. According to the above, *act*, *wrongfulness*, *fault*, *harm*, and *causation* are all required. Let us consider the requirements for *delict* individually. Firstly, we have doubted whether there is a self capable of causing an *act*, and therefore we doubt whether there is any such thing as an *act*, which is different to a mere bodily motion. We have also doubted whether there is such a thing as *moral wrong*. We have rejected the concept of *fault*, since it assumes we are causally efficacious in the inception of a morally bad event, and we have doubted causal efficacy as well as morality. We can accept that *harm* is objective and have not denied that there is such a thing as harm (eg., biological damage is harm), however, we have rejected that selves or agents can be *causally efficacious*. Therefore, if a *delict* is to be found, we have denied four out of its five requirements.

> "A person acts intentionally if his will is directed at a result which he causes while *conscious* of the wrongfulness of his conduct" (*ibid.*, p116, my italics).

On pp23-26, Neethling et al. make it clear that where a defendant is in a non-conscious state or is acting under automatism—such cases cannot yield a *delict*. Yet the implication, for example, of Libet's work, is that actions are always performed non-consciously, and hence, might be considered 'automated'.

It is quite clear that if these conclusions are true, we may have to re-think the foundations of law. For if in the face of the possible falsity of free-will, or the possibility that reactive attitudes are not justified, how we are to preserve the legal system? On p158 et seq., Pereboom discusses three methods of justification for legal systems, in order to ascertain whether they would be ruled out by hard determinism. These are: *retributivism*, *moral education or rehabilitation*, and *deterrence* (p159).

Retributivism is a system of reactive attitudes in which criminals are said to deserve punishment as a reaction for just doing wrong (p159). Retributivism is not consequentialist in rationale or effect. Retributivism has nothing (p160) to do with society's good nor anything to do with retraining the criminal. But this justification, says Pereboom, would be undermined by hard determinism, for, he claims, retributivism is merely vengefulness. Pereboom claims that no moral case can be made to justify this or differentiate it from sadism or violence that the criminals themselves perpetrate, and since it's not strictly intended to have an educative purpose, it is intended to cause further suffering (p160) for the purpose of taking revenge. So retributvisim is wrong on two counts; one: it causes further suffering for no purpose other than vengeance or the expression of the reactive attitude of outrage, and two: it is not justified if the reactive attitudes are not justified, even if punishment brings a sense of societally-sanctioned closure to the matter, rather than leaving a persistent vendetta in place (as would mere revenge) (p161).

> "Punishment as retaliation for wrong acts ought no longer to be defended in a cultivated society; for the opinion that an increase in sorrow can be made good again by further sorrow, is altogether barbarous." (Schlick, 1939, in Double, p80. See also Double, p224).

Many writers maintain that social order systems can be justified as behavioural modification or control mechanisms, rather than as moral retribution systems.[120] Pereboom refers to this as the *moral educative* model (p162 et seq.). This justification for legal systems is compatible with hard determinism because its purpose is to train the wrongdoer to behave better in future so he can become acquainted with the benefits of good behaviour, and to educate him as to the seriousness and consequences of wrongdoing,[121] as well as the suffering he inflicted on his victims (p163).

Quarantining (ie., detaining or imprisoning) could also be compatible with hard determinism as well, as long as it is construed purely as a quarantine, isolating a problematic person so they cannot persist in causing problems (p174). The extent of the detention would be in proportion to how long the criminal is likely to persist in being a threat. The person would be treated as if diseased, for the benefit of himself (to heal) and society (so the disease cannot perpetuate) (p178). Hampton (in Pereboom) suggests that the offender could instead be rehabilitated by restricting his freedom and requiring of him that he performs some community service which is

[120] see eg., G. Strawson, P. F. Strawson (p4), B.F. Skinner, J. J. Smart, and M. Schlick, *Problems of Ethics*.

[121] By "wrongdoing" here, I mean *material objective harm*, not *morally incorrect behaviour*.

appropriate given his offence (p164). This is the sort of view I wish to advocate. The legalities could remain; it would merely be their theoretical underpinning that would need to change—and perhaps some of the practical implementation of the methods used by the penal systems. For example, the assumption of our legal systems is that force (ie., punishment) is the best way to correct behaviour. This may not be the case (Skinner, p34). Pereboom describes using behaviourist conditioning techniques to prevent repetition of the bad behaviour—ie., Deconditioning (p180-182).

Another example of a possible new approach is cognitive-behavioural therapy. Criminals try to find excuses to justify their acts. Cognitive-behavioural therapy would aim to to change this process of how criminals conceptualise their acts (p183) and thus help the criminals change their behaviour patterns. Dennett (in Pereboom) says that even if determinism renders us not responsible, the practice of treating persons as if they were responsible may actually make them control themselves better. And evidence cited in Pereboom seems to confirm this (p184). We can provide people with other determinants which can lead to behavioural changes (p185).

Finally, we turn to *deterrence* theories of legal justification (Pereboom, p166 et seq.). The theory behind this model is that fear of punishment prevents crime. This is also compatible with hard determinism. But there are some concerns, eg., deterrence may require inappropriately severe punishment to ensure no-one else dares do the same thing. Similarly, if we couldn't find the perpetrator, we could, on this model, justifiably frame and punish an innocent person, since the purpose of law would be just to make an example of someone (p167). If, however, deterrence is construed as a right to harm another in self-defence or defence generally (p188), then it is also compatible with hard determinism and doesn't entail the possibility of excessively severe punishments; except where it derives implicitly from retributivist thinking (p172). If deterrence is indeed based on retributivism, then hard determinism is incompatible with deterrence theories. But deterrence is not necessarily incompatible with hard determinism; it would stand as a form of behavioural reinforcement which conditions people to adjust their behaviour—it just depends on how we justify the deterrence.

If my considerations in this section are plausible, we can see that behaviourism—the shaping of persons' behaviours for the purpose of modification of the behaviours into desirable ones—is by itself adequate as a justification for the retention of the legal system. Thus we have seen that although free-will would matter because it can justify the legal system, we can find other reasons to justify the legal system which are compatible with our previous evidence that there is no such thing as free-will. Although free-will would matter if it preserved and justified the legal system, we have seen in this section that a debatably more humane legal system could be built on top of a set of arguments which do not involve our having free-will. Thus, in the case of Law, we need not be too perturbed by our lacking free-will.

1.4. Credit for what we do, just deserts

Just deserts

There is a concern that if people lacked free-will, they would never be entitled to their "just deserts", since prior circumstantial determinants would cause whatever they did. Thus if free-will does provide us with entitlement to our just deserts, we would lose something that mattered to us if we lacked free-will. But if our 'just deserts' are conceived of as species of the reactive attitudes, in some cases, then it's not that obvious that free-will *would* guarantee us our 'just deserts' anyway. Refer back to the section on the reactive attitudes to see the argument for this.

Furthermore, suppose there is no free-will. If this were the situation, no-one could claim to be the one true author and cause of an achievement or evil deed, because their actions would be explained by prior circumstantial determinants. Thus we would lose the moral sense of someone's 'just deserts', where people 'deserve' something with moral properties as a consequence of what they do. This may seem a matter of great concern.

If my considerations thus far on the nature of responsibility have been reasonable, we would just be left with the causal sense of 'desert'—in which our actions beget a reaction, and that reaction is in itself not so much a reward or punishment, but merely a consequence. Since I have argued that there is no free-will, I now need to suggest how we can reconstruct the notion of just deserts. For if we need free-will to deserve something, and if we value being entitled to something, then free-will would matter. So I suggest that instead of people being "rewarded" or "getting what they deserve", we now have to just think of people "needing encouragement" or "needing training". Again, we could retain some of our practices by just changing their rationalisation. Harm, for example, would beget a natural reaction of violence, without any particular moral significance. In other words, we can reconceptualise 'desert' as referring to a reaction which is appropriate in view of the natural causal sequence, no longer having moral properties. Similarly, rewards or credits would merely serve as markers of what a particular human is capable of, instead of being morally-warranted forms of praise.

Admittedly, one can desire to be praised, but one might not want to be encouraged or trained. This does indicate, of course, that there is a difference between what we normally do (praise for its own sake) and what I am suggesting here (praise as an encouragement tool). But I do not think we strictly need to change our practices; all that's being suggested here is that we think differently about the reason behind performing the act of praising or rewarding, blaming, etc.

Creativity and credit

If we had free-will, then we would at least potentially be *creditable* with being the originator of some novelty; we could meaningfully be said to be the cause of some artefact or novel idea. And

if we lack free-will, we may lose the ability to be creative, and thus, free-will would matter, because we value being creative. But even if we lack free-will, I do not think the picture is that bleak. There are two aspects to the concept of creativity: the concept of *origination* (authorship), and the concept of *originality* (novelty). If we talk of someone having done something creative, what we mean is that she created something of which she alone is the author or originator, and, that she did something novel or original. Only one of these senses of creativity falls away if we lack free-will.

If what we do is predetermined by causal factors beyond our control, what we do could never originate with us as persons, and thus we could not be creative in the sense that we authored what we did. In other words, because we lack ultimate responsibility for what we do, if we lack free-will, it would never be the case that we could truly be said to be the authors of something creative. But that we *are* creative suggests that either we have free-will and it allows us to be creative, or, that we lack free-will and misunderstand the nature of creativity. I take the latter answer. We are creative because of the complex mesh of ideas and emotions we have, and these are causally necessitated historical products. So we can still be creative, in the sense of being *able* to produce something novel. What I am arguing is that what is important in creativity is our ability to be creative, ie., to generate novelties, new ideas, etc. What we lose in creativity, if we lack free-will, is the authorship of what we create. To some, this is important, because we want to be proud of what we have created. But if hard determinism is true, and if the mental is only a product of prior circumstance, what we achieve we owe to our circumstances. We now need to be humble and acknowledge the role of pure luck in our life courses, instead of bragging that we are self-made. Praise and medals, as we have said, need to now be seen rather as demarcators of capability, as signs to the outer world of what we can be useful for, rather than as rewards for our own enjoyment. This does not mean that praise and medals need to go away, just that they need to be justified on different grounds. Praise and credit, as we think of them now, are some of the "just deserts". If we have denied that people have any desert other than the causal consequence of their actions, we have not really lost the ability to reward people; reward can be a causal consequence of an action rather than a moral response; compare "remuneration" or "salary" to "prizemoney" or "winnings" to see the kind of conceptual difference I have in mind (earnings are more of a consequence of action, rather than a chance event).

1.5. Life hopes, Personal efficacy, Reasons for acting, Dignity, etc.

Life hopes and Personal Efficacy, Reasons for acting

In this section, I will argue from two different perspectives that we can retain either the causal efficacy of reasons, even in the absence of free-will, or that if we understand the reasons that we have, differently, we also would not require free-will to retain some sense of our having reasons for action. The first position is Pereboom's; the second is mine.

One might be tempted to think that if one is a product of circumstance, that one's future actions would be a product of circumstance. This is true. However, the world we live in is statistical and chaotic (in the mathematical sense)—and thus, we cannot predict what will happen to us just by looking at our individual pasts. We need to also consider our future circumstances, which we do not know about—and those future circumstances are what will shape us toward further futures. Even if we cannot guarantee in advance what will happen when we "do" something, because we cannot know all the factors that come into play, at least we know that what happens isn't *spurious*. The conscious self, given my evidence, is not really what is controlling us, or making us choose the things we seem to. But nonetheless we do feel as if we can get what we want out of life. This is because our satisfaction of our desires still happens, and it happens as a result of our bodies moving to achieve our goals, even if other factors come into play. That ought to be enough. The fact is that *this is how we are*, this is how our actions occur. But just because this is how we are, learning that we are determined shouldn't make us *feel* any less able to hope that we can get what we want out of life. We feel that we *can* get what we want out of life, and we often *do* get what we want out of life. So, if my evidence that we lack free-will is correct, it is not free-will that enables us to get what we want out of life. It is a combination of our pasts and events around us that lead to what it is that we want and how we get programmed, and thus what our bodies do. That an achievement lacks praiseworthiness does not mean it is not an achievement we did not want, or we did not get what we wanted (Pereboom, p197). It is not that we lose our achievements, we just learn gratitude for those who supported, raised, and taught us, for our luck and opportunities. We learn humility.

Pereboom asks (p187): Would we lack control over our lives and resign ourselves to fate if hard determinism[122] were true? The answer that Pereboom ventures is *No* (p188). He maintains that accepting this model of our choices would make us more accepting and serene towards how we are treated and what happens. We would have to treat each other more calmly, and thereby reduce anger and resentment—which cause most of the misery we experience. A further advantage to this view is that it allows one to forgive oneself for failure—because one can now recognise that many circumstances and determinants are not under one's control—even aspects of who we are as persons (p193). Our life hopes depend on us being able to get what we want out of life—because we wanted it, ie., because of our reasons.

Pereboom (pp134-7) discusses the extent to which our deliberations about our reasons for acting, could make a difference. I have argued that we are not agent-causes (as does Pereboom). But it does feel, when we deliberate, as if there are choices that are open to us. And yet Pereboom grants that if determinism is true then we don't really have two choices; we will only take just that one which was predetermined. But the decision-making process (however one might construe this) is not irrelevant. It is part of what leads to the eventual decision; even if in advance it is determined, and even though we don't know what it will be, and even though it is inevitable, it is surely better than not deliberating at all (p138). Nagel (in Pereboom) expresses a concern that if everything is inevitable, we won't bother to try accomplish anything, since it's a

[122] or in his case, hard incompatibilism

foregone conclusion whichever way the world goes, ie., our deliberations and actions may turn out to have no effect since it is all antecedently determined. But, Pereboom counters, antecedent determination of our choices and actions does not undermine their causal efficacy. In other words, just because we're determined by antecedent factors, it doesn't mean that we ought to not bother about making decisions, and throw our hands up in despair. If the world was determined, it would not mean that we would know in advance how everything would go, and thus resign ourselves to fate. The world is not pre-ordained, and our decisions do make a difference, even if they are antecedently determined. Because we cannot say in advance *what* we will decide. The decision-making process is part of the deterministic sequence of the world. And what we decide will have an effect. Though this does not make us free or accountable, it does return to us a sense of there being a point to deciding. Similarly, Pereboom argues, we would not suddenly be unresponsive to reasons; reasons can be determinants as well (p138, p139). We could hold people rationally accountable; ie, demand of agents they adhere to reason, justified as a way of training people to be more rational, and that this will affect their behaviour. We are capable of modifying our behaviour in the light of reasons, and achieving better things by being more reasonable. This is not to say that we are free, or have a choice, or are deserving of reactive attitudes. Just that we can achieve better results if we think our actions through. This argument is not incompatible with our defence of hard determinism (Pereboom, p139), though the evidence from Libet does raise some questions. Pereboom is thus arguing that reasons for acting, and the decision-making process, are deterministic, and lead to us achieving what we want. He also acknowledges that it is not up to us as to what reasons there are, and thus we are determined and unfree because we are led to do what we do because of the reasons that there are. I believe that this characterisation is acceptable to a hard determinist. The primary difference, then, between my position and Pereboom's, is that Libet's evidence seems to indicate that the reasons, and reasoning process generally, would all be effected non-consciously, and the reasons and desires themselves would not be causally efficacious. Thus we now consider my view, since the evidence from Libet seems to present a problem for Pereboom's view.

We have seen that in order for persons to be able to act for reasons, a causal link is required between an agent and her actions. Of course, if some of the evidence presented in this paper is true, then persons might not be efficacious *qua* persons. What we call "the choices of a self", if these are merely non-conscious brain events, needs to be re-characterised. If the research presented in this paper is correct, this conscious collective of mental states (the "self") is merely epiphenomenal on preceding non-conscious states. Thus we have to re-think what we understand by causal efficacy of persons, and the extent to which we can impact the world. Consider the old idea that there was a substance called "caloric" which, when it passed into an object, was the reason the object was hot. In other words, caloric itself was heat. But with further research, scientists now know that what we thought was caloric, was actually just molecular energy (motions). Caloric itself does not exist. Let us now apply this analogy to the causal self. What we thought was X (eg., a causally efficacious self), was actually just the effect of a mechanism Y (a non-conscious set of brain states), which we now understand as a mere mechanism rather than a thing. Mental states, which we thought to be causally efficacious things,

are not in fact causal at all. They, and our bodily motions, are both effects *of* the same things: the non-conscious brain states. What we call "persons acting" is actually just a shorthand for many complex brain processes and their effects, just as "heat" is shorthand for the processes of molecular agitation. But just because we have an explanation for the mechanism behind heat, it doesn't mean we now have to use elaborate descriptions to refer to that mechanism; we can still say 'it is hot' as a shorthand. The same goes for the self. Just because we now have an explanation for how our choices work, it doesn't mean that the choices will now feel unsatisfactory, or puppetteered. We just have a new explanation for what a 'choice' is; it's just nothing like what we thought it was. True, on my model, I have dismissed the notion of a centralised self, the concept of "I" or the "ego" or "self", as with "caloric". But this does not mean that the subjective experiencing goes away. Similarly, because we now have an explanation for how choices occur non-consciously, it doesn't mean that we now have to say some kind of choosing does not occur, or that choices are now somehow unsatisfactory. Perhaps we no longer "act" for reasons but merely "behave" because of non-conscious causes. But it is at least clear that we do not need free-will for our desires to be satisfied. Our desires and our satisfying of them is carried out by a non-conscious brain state, not by our conscious "reasons". A "reason" is generally described as being a *desire* for something, and being able to *see that* doing something will obtain that object of desire, and it involves *believing* that doing something will obtain that object of desire. If I have a reason to do Y, as a human, it means that I can see that doing Y will be beneficial. We say that an agent *A* acted to obtain Y, and *A* thought that acting in a certain way would attain Y, and *A* desired Y—her reason for acting was to obtain Y.

So the question is: What matters to us: does free-will matter to us because it will allow us to act for reasons, and get what we want because of reasons that we have? Or, would it just be enough that we get what we want because of a brain state originating within us, that makes us go get what we want? From a very idiosyncratic perspective, the latter diminished prospect doesn't bother me. I recognise that it is not the "reasons" (ie., the mental states of desire and belief) that will motivate us. That doesn't bother me, for three reasons: firstly; I still have those mental states and enjoy them. Secondly, I get what I want anyway. Thirdly, if Libet's evidence is correct, then this is a true representation of how we achieve anything or get anything we want, anyway, and there's no point to being upset about how the world is. Our "conscious" "mulling" or "rumination" would just be a mental event consequent on some antecedent brain event. Our conscious reasons or ruminations, certainly, would not be what cause us to act. But they would all have antecedent non-conscious brain events, and those brain events would most likely be the results of some or other non-conscious computational process which took in evidence from the environment and senses, and caused the body to act to carry out the result of that computation. I believe this is really what causes our experience of the reasoning process, even if it is not the *conscious* reasoning which makes us do what we do.

One might argue that a benevolent dictator could give us what we want, and yet we would not feel as if we were satisfied that it was we who achieved it ourselves. This is true, but phenomenologically, that's not how we experience it. I am acknowledging that it is a brute fact

that we have a phenomenological experience of satisfying actions. This phenomenological experience, in the light of our lacking free-will, needs explanation. The explanation I am offering is as follows. There are internal non-conscious states of privation, which lead to our our phenomenological experience as of having a desire for something. There are internal non-conscious systems of computation, which lead to our phenomenological experiencing as of having a reasoning process. And there are internal non-conscious systems of causation, which lead to our phenomenological experiencing as of choosing to act. Our experiencing of efficacy at the subjective phenomenological level depends on an underlying causal deterministic mechanism, which not only ensures we experience desires at all, but that we do in fact get what we desire.

In the science of Biology, we characterise what a system does in terms of its function, so, for example, a heart has a function of pumping blood, phototropism in sunflowers tracks the light source, etc. My model of human behaviour is similar; just as the sunflower tracks the sun for the reason of absorbing light, so humans perform certain actions for certain functional reasons. But this is not the same as saying that their conscious reasons—those mental states—cause the behaviours that we observe. I recognise that this does not rescue the causal efficacy of our mental states—what we call 'our reasons'. But since we still get what we want out of life, to a large extent, I don't see that we have reason to mourn this loss.

Distinction from the animals, human dignity

Some writers might argue that free-will is, like language and culture, one of the hallmarks of being truly human. It may be argued that animals, unlike humans, are Frankfurt-wanton. Free-will may be a way to make good on the difference between humans and animals. But since we have denied free-will, we have to demarcate our species using other factors, like culture, language with grammatical structure, etc.

> "We do not fall from the grace of personhood simply because the theoretical prize of freedom turns out to be illusory" (Double, p225).

Individuality

There is a concern that free-will, as an expression of the individual will, preserves individualism and the uniqueness or specialness of every person. But individuality is guaranteed by unique life experiences (regardless of where those life experiences come from)—and they create a unique person from their circumstances. Individuality is preserved, regardless of free-will.

Summary

Given these considerations and my attempts to reconstruct the moral, cultural and interpersonal desiderata of our lives, it may not seem as hopeless as one might first think. Even on these crude reconstructions, it is apparent that we could retain some of them even in the face of not having free-will. Whether free-will matters is thus a question of which desiderata I have failed to preserve, and the extent to which these things *do* matter to us. I think, for example, that law does matter, and that I have preserved it. Whereas self-control and reasons for acting are more dubious—I believe more work (within the context of Libet's evidence), would be needed to preserve these desiderata. In some cases it is clear that they are not warranted. In such circumstances, we have to modify our view on the desideratum which we have lost. We have to not only re-evaluate it, but also perhaps substitute it with another attitude which is justified in the context of this framework. Consider blame. It assumes a person knowingly did wrong. If this reactive attitude requires free-will, we have lost the justification for it. We could, perhaps, replace it by pity, and train ourselves to pity people who do wrong. For pity does not assume a person is free; pity is inspired by our recognising someone as behaviourally abnormal, or down on her luck. A person who did wrong, presumably, would just be someone who had had harmful prior determinants; she was unlucky.

Section 2—Why we believe in free-will—Some speculations

If we apparently lack free-will, it is quite mysterious then, as to why we should persist in believing in it. Perhaps it is a necessary illusion—so we can have self-respect. Perhaps we need to continue to believe that we can and do get what we want (Smilansky in Pereboom, p131). I know I do. People can also persist in false beliefs even if they are aware of contrary evidence. So perhaps it is not surprising that people will persist in such a belief. This sections is, however, not intended as a detailed exploration, but merely some other speculations as to why people persist in the belief in free-will. I suspect there are a number of reasons, such as:

a. Religion teaches that man is endowed with free-will.

b. Criminal penal systems and reward or recognition systems, such as prizes for winning competitions, seem to require free-will and have existed for millenia. We want to be able to praise people who do right, and punish people who do wrong.

c. We feel causally efficacious, and able to just do as we please, and because we feel as though we can do as we please, we assume we are free to do as we please.

d. This is highly speculative, but I suspect we infer accountability and agency because of our experiencing of reactive attitudes; *because* we react to others, we assume they are accountable.

These speculations above may explain why people do believe in the concept of free-will—it is because of an old innate belief of some kind, a mistaken interpretation of our inner life. We *think* we are one person, we *think* what we "choose" is chosen by us alone, spontaneously, and that whatever we choose to do, does in fact happen because we chose it. But if my research proves correct, this is all *false*. Consider, by analogy, the impression that the world is flat, or the impression that the sky is a solid blue dome, etc. Though substantial problems for these beliefs exist (as we have seen), the impressions persist. That's just the way our brains are—for evolutionary reasons. Animals that had the ability to recognise other animals as agents with intentions and assumed that other animals had a "will", are animals that have survived (cf. Dennett, *The Intentional Stance*). More primitive societies have believed in animism—that is, that ordinary objects have "souls" and are capable of doing things—ie., have "powers". Since science has stripped all objects in the world—except persons—of such animistic properties, and reduced all objects to automata, people are reluctant to bite the bullet, and accept that perhaps people are also just automata, albeit more complex than any others. It is my observation that persons, for example, might talk to inanimate objects, or hit them, or whatever, but if pressed, would deny the object had any agency. This 'intentional stance', as Dennett puts it, persists in regards to ourselves. Thus, while we no longer believe the volcano goddess Pele has it in for us when she erupts, we still believe our fellows 'have it in for us'. We still react to inanimate objects, and we still react to persons. But that means nothing about their having free-will. Yet we still impute free-will to persons—presumably because we still feel entitled to the reactive attitudes where persons are concerned.

What worries me, however, about the idea of free-will is its darker uses (as mentioned above in point *b*.). As Glover points out, blame is often a tool of control (p462)—but we can get the control (eg., law)—without requiring blame. Consider this example from Watson.

> "I happened to share a table some time ago with a high-ranking official ... During the conversation, his companion bemoaned her unsuccessful campaign to get the City Council ... to take action to ameliorate the plight of homeless people in [the] city. With a sense of bewilderment ... she noted that [the Council] had been willing to fund shelters for dogs and cats while it did nothing for people. The official was untroubled [by her complaints]. 'The difference', he said, 'is that people have free-will'." (Watson, 1987, p161).

I realise that not all people want free-will *just* so they can blame, but that it may at least in part derive from this desire. Free-will may be a rationalisation for what are actually instincts: the reactive attitudes of blame, resentment and anger. Pereboom says (p210) that we persist in believing in free-will so we can justify our continued recourse to these emotions, justified by the claim that the person who receives the angry response 'deserves' it. Consider these remarks:

> "Everywhere accountability is sought, it is usually the instinct for *punishing and judging* which seeks it" (Nietzsche, *Twilight of the Idols*, p64).

> "No doubt there are some atavistic and bloodthirsty elements in our current vision of responsibility, dignity and guilt. ... One may speculate with Nietzsche about dark undercurrents of violence in our psyches" (Dennett, 1984, p154, p159)

> "Why do we want so much to hold others responsible? Could it be a streak of sheer vindictiveness or vengefulness in us, rationalised and made presentable in civilized company by a gloss of moral doctrine?" (Dennett, 1984, in Double, p81).

> "... it is significant that anthropologists have suggested that most human societies can be classified either as 'guilt cultures' or as 'shame cultures'" (G. Strawson, p9. See also Wallace, p38).

I argue that, like any other animal in nature, we operate on instincts and learnt responses. We are determined to be what we are, by circumstance and heredity. I believe that this leaves us with a metaphysically cleaner view of ourselves, ready to better deal with the human condition, and human suffering, from an objective scientific point of view, stripped of old metaphysical relics like "souls" and "wills" and "just deserts". We stand to lose little, and gain much in the way of mutual understanding and tolerance. Einstein (in Pereboom) agrees that it is not up to people as to what it is that they want to do, and argues that with this view we could be less angry and violent (pp211-212).

> "We can tell them that they can go right on talking and thinking the way they were—all they have to do is give up this bit of metaphysical baggage. ... [souls] are [like free-will] only *abstracta*" (Dennett, 1993, p367).

> "Far from threatening meaning in life, hard incompatibilism can help us achieve the conditions required for flourishing, for it can assist in releasing us from the harmful passions that contribute so much to human distress. If we did in fact relinquish our presumption of free-will and moral responsibility, then, perhaps surprisingly, our lives might well be better for it." (Pereboom, p213).

> "*The error of free-will*. We no longer have any sympathy [...] with the concept of 'free-will'... *No one* is accountable for existing at all, or being constituted as he is ... that the kind of being cannot be traced back to a *causa prima*... this alone is the great liberation... We deny accountability: only by doing *that* do we redeem the world." (Nietzsche, *Twilight of the Idols,* pp64-5).

Appendix

Baars, B., Baggot, M. and McCrone, J. (2000). "Libet replications and implications", *eScribe: PSYCHE-B: Internet.*

Libet replications and implications

* From: John McCrone * Date: Thu, 24 Aug 2000 05:06:18

[Stan Klein:] Facts: Libet showed that a thalamic stimulus requires a duration of about 250 msec to be felt whereas a 20 msec skin stimulus is adequate.

[Gilberto Gomes] Sorry, this is not a fact. Libet adjusted the intensity of the stimulus so as to have a Minimum Train Duration of 200-300 ms. With a weaker (liminal) stimulus, the requirement would be of an even longer duration (up to 2000 ms, Libet et al. 1964, p. 557), but with a stronger stimulus a duration much shorter than 250 ms would be adequate.

Libet was careful to distinguish between the type of stimulation that produced realistic sensations (a furrowing of the skin, taps, jabs, brushes and occasionally a flush of warmth or coldness) from stimulation that simply provoked a harsh pins and needle tingling. Realistic sensations needed a weak current for about half a second (with a range of third to several seconds). Strong jolts of 100 or 200 ms could produce a conscious sensation, but they were experienced as just a jolt - and Libet's other experiments, such as backward masking and forward enhancement, suggest that even the harsh jolts were not felt as they happened but had to be integrated into the stream of experience over the course of at least a third of second.

I've spent a lot of time considering Libet's experiments (they were central to the arguments in my last book, Going Inside, and I

reviewed the whole history of his work). I actually ended up feeling that they are far more inconclusive than they first appear. It seems likely that the actual integration time (the time needed to become focally aware of a sensory event) is highly dynamic and not tied to some magic half second figure. But still, the very fastest time of processing would seem to be about 300 milliseconds - if you are expecting something to happen at a location, and roughly aware of what it might be, then you can complete the attentional act relatively quickly. I have the feeling that Libet got an average of half a second because he used a weak threshold stimulus, so causing the brain to do extra attentional work. His subjects were waiting for expected events, but so as to produce realistic sensations, the current had to be turned down to a "nagging" level where their brains would be forced to make some memory-based interpretation and this stretched out the processing time somewhat.

On the other hand, in cases where we are not expecting a sensory event, and it is also rather ambiguous, then processing times probably stretch right out to a second or more. It would take much longer than half a second to get the event into the bright spotlight of attentive awareness.

But as Klein argues, whether the processing delay is 100ms, 300ms, 500ms or 1000ms, any theories about brain processing and consciousness have to treat neural processing delays as a basic fact (see for example Nunez, Paul L. (2000) Toward a Quantitative Description of Large Scale Neocortical Dynamic Function and EEG Behavioral and Brain Sciences 23 (3)). Neural traffic takes time to propagate. If top-down attentive processes have to focus activity into a coherent meaningful pattern, then the delays will be pretty long. Wundt understood all this over a century ago.

So the brain must take time to integrate a state of consciousness (if you believe that neural patterns are the key to consciousness). This means it probably does its best to compensate for the inevitable delays. And it does this first by being an anticipation-based system (predicting as much as possible in advance of it happening) and secondly by having a filter of fast, unthinking, habits that can intercept events at a preconscious level, dealing with them as learnt routines (as when we drive cars or climb stairs). There is abundant evidence to show that this habit level of processing takes only100 to 200ms to complete.

Libet's claims have to be judged against a backdrop of other evidence. For example, the psychophysical phenomena of iconic memory and the attentional blink both bolster the feeling that half a second is a reasonable benchmark for attention forming processes in the brain. Iconic memory can be interpreted as a delayed "shrink" of the perceptual field. Instead of creating an escalated focus as fast as possible, the escalation is deliberately delayed, so telling something about the natural cycle time of the process. In a typical experiment, a slide containing three rows of four letters is flashed up on a screen for just 50 milliseconds. Subjects have to avoid reading any of the letters until a tone sounds to tell them which column they are mentally suppose to turn to and report. As long as the signal is not delayed more than a third to half a second, their performance is almost perfect. They can inspect a still lingering iconic image and read off the chosen column. But this act of looking then wipes out all memory for the other two columns, as if the top-down act of focusing irrevocably sculpts what had been a flat and even spread of preconscious mapping. See "The information available in brief visual presentations," G Sperling, Psychological Monographs 74:11 (1960). Note that it has been suggested that the cortex grabs the information direct from lingering retinal activity. See "Locus of short-term visual storage," B Sakitt, Science 190, p1318-1319 (1975).

The attentional blink is a more recently discovered effect which reveals the brain needs about half a second to recover from being "pinched up" to catch a perceptual event-it cannot focus sharply on a second event until it has had time to realign its anticipatory state. In a typical experiment, subjects are shown a series of alphabet letters and asked not only to report a sighting of any x's, but also the occasional insertion of the number 4. When asked to spot x's or 4's alone, people can catch them all even at a presentation rate of eight a second. But with two clashing targets to report, a half second mental blindspot appears. This suggests that the more rapid attentional performance is only possible because a single anticipatory framework remains in place. Once subjects have to swap between goal states, the full cycle time is exposed. See "Temporary suppression of visual processing with an RSVP task: an attentional blink?" J Raymond, K Shapiro and KM Arnell, Journal of Experimental Psychology: Human Perception and Performance 18, p849-860 (1992). For evidence that information is still being handled at

a preconscious level during the blink, see "Word meanings can be accessed but not reported during the attentional blink," SJ Luck, EK Vogel and KL Shapiro, Nature 383, p616-618 (1996).

The P300 literature, the temporal characteristics of NMDA channels, and much more could be wheeled in to support the idea that attentional states (the time it takes to escalate a sensory event to bright awareness) takes a surprising time.

Finally, in answer to Gilberto, there have been some attempts to replicate Libet's work. For example, check recent papers by Kimford J. Meador, at the Department of Neurology Medical College of Georgia.

I include here a summary that Meador sent me of his results.

[Kimford Meador] 1. Reproduction of Libet's finding that perceptual threshold drops as the train duration is extended out to about 300ms. This is true for central brain electrode stimulation in patients, transcranial magnetic stimulation in healthy volunteers, and even for electrical stimulation to the hands. 2. This train duration effect is markedly enhanced in patients with neglect syndrome. I believe that this enhanced effect is telling us something special about the disorder of neglect as we have shown that the effect is not due to simple defects in primary sensory thresholds or in vigilance. 3. Reproduction of Libet's finding that an evoked potential can be recorded from the primary somatosensory region for stimuli below perceptual threshold. 4. Demonstration that the perceptual threshold is lower in the left than right hand of dextral (right handed) people across a large age range. In contrast, sinestrals (left handers) have equal left/right thresholds. The lower left hand threshold in dextrals exists on a central basis although there may be a matched peripheral component. This left/right asymmetry in perception in people with typical cerebral lateralization (in this case dextrals) is postulated to be due to a right brain dominance for directed external attention. It is not present or less marked in sinestrals as a group because they tend to have such functions more evenly distributed across both cerebral hemispheres. 5. Demonstration in two separate cohorts that a masking stimulus (applied to a hand contralateral to a hand receiving the target stimulus) has its maximal effect when applied at 50-100ms after the target stimulus. This is clear evidence that

conscious perception is delay and that the mechanisms of conscious perception are particularly vulnerable at this time period. Our present work involves the role of gamma activity in these processes. This work is going well but is still just a bit too early for publication. Our basic view of the the brain mechanisms involved is that processing is initially subconscious. Later the results of that processing may gain access to conscious awareness. If the stimuli are weak, this may require summation as we see with the train duration effect. However, the summation only occurs over a few hundred milliseconds as longer train durations have no additional effect. It is unclear to me if the actual conscious perception is delayed several hundred milliseconds, but it is clear that it is certainly delayed 100ms. What then is the mechanism involved in the actual conscious perception? We believe that it involves coherence (i.e., synchronous) neural activity in the gamma (i.e., 30 Hz) range between the thalamus (and possibly the upper midbrain) and the cortical areas involved in the processing of the specific stimulus which is perceived. The cortical area might be the primary sensory area for simple stimuli in the specific modality, but might involve higher order regions for more complex perceptions (e.g., fusiform gyrus for facial recognition). In the case of masking, we think that this process is truncated at the thalamus by the masking stimulus which arrives 50-100ms later.

Meador's refs (sorry, I've only got the preprints)

PHYSIOLOGY OF SOMATOSENSORY PERCEPTION: CEREBRAL LATERALIZATION AND EXTINCTION by KJ Meador, PG Ray, L Day, H Ghelani, DW Loring.

TRAIN DURATION EFFECTS ON PERCEPTION: SENSORY DEFICIT, NEGLECT, AND CEREBRAL LATERALIZATION by K.J. Meador, P.G. Ray, L.J. Day, D.W. Loring.

I too am off for the next few weeks so apologies to anyone who wants to take up specific points.

from John McCrone

check out my consciousness web site
http://www.btinternet.com/~neuronaut/

Re: "backwards referral"

* From: Michael Baggot * Date: Thu, 24 Aug 2000 05:11:20

Here is a thought I had while perusing a paper on Libet and sensory delay by Ted Honderich on the web. I am merely throwing this out here with an invitation to comment as I am only vaguely familiar with Libet's work although I have heard him speak.

Has anyone considered the possibility that the AEP (average evoked potential) occurring in an already experienced test subject at very short latency and well in advance of actual conscious awareness triggers a preprogrammed motor response such that the subject would not in fact be responding to a conscious formulation of the stimulus but to this much more appropriately timed *trigger signal*? It seems to me that this kind of habituated response would eliminate any need for backward referral both in a test situation and in real life situations like hitting a tennis ball or baseball.

Michael Baggot baggot@cruzio.com

Re: Libet replications and implications

* From: Bernard Baars * Date: Thu, 24 Aug 2000 16:04:28

Good review of the temporal delay effects by John McCrone. I just want to point out that a half second delay in a critical decision can kill you on the Serengeti plains, whether you're a predator or prey. (A predator can die by being deprived of food.) It is well known that the answer is to make things predictable, and that if you can predict an event with temporal accuracy you can routinely obtain NEGATIVE reaction times. That is, you can react to the stimulus before it comes. That doesn't mean it takes no time for the stimulus to come to consciousness, of course, but only that you can act on a

highly predictable stimulus in time to catch it in the act. Without that animals would starve. (That's one explanation, by the way, why lying in ambush is such an effective predation strategy. The predator in ambush can predict exactly when, where, and how to jump the prey, while the unwary prey has to pay the temporal price of unpredicted reaction time.)

On a related issue, it was said for decades that Pavlov's optimum delay for conditional association of CS to UCS was 2.5 seconds. That's what Pavlov found, and the learning efficiency dropped off quickly for longer delays. That is utterly absurd. It implies that animals can only learn to anticipate events by 2.5 seconds, a time horizon that would get you quickly dead in any natural situation. Fortunately John Garcia came along to show that wolves (and others) were able to learn to identify a poisoned food for a novel-tasting food with a CS-UCS interval of 24 hours or more. So it appears that animals can learn conditional associations at much longer time intervals than the classical Pavlov interval, providing there is some biologically plausible relationship, such as food being toxic and getting sick. In nature there probably are no totally arbitrary associations, like a bell signalling meat powder in Pavlov's lab.

Just an ecological side note.

Best,

Bernie Baars

Re: Libet replications and implications

* From: Stan Klein * Date: Thu, 24 Aug 2000 21:44:05

[Stan Klein:] Facts: Libet showed that a thalamic stimulus requires a duration of about 250 msec to be felt whereas a 20 msec skin stimulus is adequate.

[Gilberto Gomes] Sorry, this is not a fact. Libet adjusted the intensity of the stimulus so as to have a Minimum Train Duration of 200-300 ms. With a weaker (liminal) stimulus, the requirement would

be of an even longer duration (up to 2000 ms, Libet et al. 1964, p. 557), but with a stronger stimulus a duration much shorter than 250 ms would be adequate.

Duration of a threshold stimulus of course depends on stimulus intensity. I used the 250 (or 300) msec number because it is my recollection that in almost all of his backwards referral experiment that is the duration he used. So in an effort to get to the essence of the issue that's the number I quoted. The problem with some of the analyses of Libet is that they make it look like the details were complicated. But they actually aren't that bad. However, I do believe I agree with McCrone that Libet's data is sloppier than people often realize. Glynn made that point a long time ago and it was the main point of my plots in the Tucson III Proceedings.

References

Aronson, J., Dietrich, E., Way, E. (1992). "Throwing the conscious baby out with the Cartesian bath water" in *Behavioural and Brain Sciences,* 15:2

Baars, B. (2000). "Libet replications and implications", *eScribe: PSYCHE-B: Internet.* Attached as an appendix.

Baars, B. J., Fehling, M. (1992). "Consciousness is associated with central as well as distributed processes" in *Behavioural and Brain Sciences,* 15:2

Baggot, M. "Libet replications and implications", *eScribe: PSYCHE-B: Internet.* Attached as an appendix.

Breitmeyer, B. G. (1985). "Problems with the psychophysics of intention" in *Behavioural and Brain Sciences,* 8:4

Block, N. (1980). *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard.

Block, N. (1992). "Begging the question against phenomenal consciousness" in *Behavioural and Brain Sciences,* 15:2

Chisholm, R. (1976), in **Double, R.** (1991). *The Nonreality of free-will*. Oxford: Oxford

Chisholm, R. (2002). "Human Freedom and the Self" in Kane, R. (Ed.) (2002). *Free-will.* Oxford: Blackwell.

Danto, A. C. (1985). "Consciousness and motor control" in *Behavioural and Brain Sciences,* 8:4

Davidson, D. (1963). "Actions, Reasons and Causes" in *The Journal of Philosophy*, Volume LX, No 23.

Dennett, D. C. (1984). *Elbow Room*. Oxford: Oxford

Dennett, D. C. (1993). *Consciousness Explained*. London: Penguin.

Dennett, D. C., and Kinsbourne, M. (1992). "Consciousness and the observer: The where and when of consciousness in the brain" in *Behavioural and Brain Sciences,* 15:2.

Dennett, D. C. and Kinsbourne, M. (1992). "Escape from the Cartesian Theater" in *Behavioural and Brain Sciences*, 15:2

Dennett, D. C. (2002). "I Could Not Have Done Otherwise—So What?" in Kane, R. (Ed.) (2002). *Free-will*. Oxford: Blackwell.

Descartes, R., *Discourse on Method*, London: Penguin, also available on Internet

Doty, R. W. (1985). "The time course of conscious processing: Vetoes by the uninformed?" in *Behavioural and Brain Sciences,* 8:4

Double, R. (1991). *The Nonreality of free-will*. Oxford: Oxford

Eccles, J. C. (1985). "Mental Summation: The timing of voluntary intentions by cortical activity" in *Behavioural and Brain Sciences,* 8:4

Fischer, J. M. (1994). *The Metaphysics of Free-will*. Oxford: Blackwell

Frankfurt, H. (1969). "Alternative Possibilities and Moral Responsibility" in *Journal of Philosophy*, vol 45

Frankfurt, H. (1982). "Freedom of the Will and the Concept of a Person", in *Journal of Philosophy*, vol. LXVIII, No. 1, Jan. 1971, reprinted in **Watson, G.** (Ed.) (1982). *Free-will*. Oxford: Oxford
Giancoli, D. C. (1991). *Physics: Principles with Applications*. USA: Prentice-Hall.
Ginet, C. (2002). "Freedom, Responsibility and Agency" in Kane, R. (Ed.) (2002). *Free-will*. Oxford: Blackwell.
Glover, J. (1983). "Self Creation" in *Proceedings of the British Academy,* vol LXIX
Grossman, R. (1984). *Phenomenology and Existentialism*. London: Routledge.
Honderich, T. (1973) (Ed.). *Essays on Freedom of Action*. London: Routledge.
Honderich, T. (1993). *How Free Are You?*. Oxford: Oxford
Honderich, T. (1984). "Is The Mind Ahead Of The Brain?—Benjamin Libet's Evidence Examined" Internet: *http://www.ucl.ac.uk/~uctytho/libet1.htm*. Originally published as "The Time of a Conscious Sensory Experience and Mind-Brain Theories", Journal of Theoretical Biology (1984) 110
Honderich, T. (1986). "Is The Mind Ahead Of The Brain?—Rejoinder To Benjamin Libet". Internet: *http://www.ucl.ac.uk/~uctytho/libet2.htm*. Originally published as "Mind, Brain and Time: Rejoinder to Libet", *Journal of Theoretical Biology* (1986) 118
Hoffman, R. E., and Kravitz, R. E. (1987). "Feedforward action regulation and the experience of will" in *Behavioural and Brain Sciences*, 10:4
Hume, D. (1748). *An Enquiry Concerning Human Understanding*. London: Penguin, also available on Internet.
Hume, D. (1740). *A Treatise of Human Nature*. London: Penguin, also available on Internet.
Jacquette, D. (1994). *Philosophy of Mind*. Prentice-Hall: New Jersey
Jeannerod, M. (1992). "The where in the brain determines the when in the mind" in *Behavioural and Brain Sciences,* 15:2
Jung, R. (1985). "Voluntary intention and conscious selection in complex learned action" in *Behavioural and Brain Sciences,* 8:4
Kane, R. (1996). *The Significance of free-will*. Oxford: Oxdord
Kane, R. (2002) (Ed.). *Free-will*. Oxford: Blackwell.
Kim, J. (1989). "The myth of nonreductive materialism", in *Proceedings of the American Philosophical Association*, Vol 63 No. 3.
Latto, R. (1985). "Consciousness as an experimental variable: Problems of definition, practice, and interpretation" in *Behavioural and Brain Sciences,* 8:4
Leon, M. (1997). "On the Value and Scope of Freedom". Pre-publication edition, available in *Ratio*, vol. 12. (1999)
Leon, M. (2001). "The Willing Addict: Actor or (helpless) Bystander?" in *Philosophia*, Volume 28
Leon, M. (2002a). "Responsible Believers" in *The Monist*, vol. 85, no. 3
Leon, M. (2002b). Personal communication.
Lewis, D. (1978). "Mad Pain and Martian Pain", in **Block, N.** (1980). *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard.

Libet, B. (1982). "Brain stimulation in the study of neuronal functions for conscious sensory experiences" in *Human Neurobiology* 1
Libet, B. (1985). "Unconscious cerebral initiative and the role of conscious will in voluntary action" in *Behavioural and Brain Sciences,* 8:4
Libet, B. (1985). "Theory and evidence relating cerebral processes to conscious will" in *Behavioural and Brain Sciences,* 8:4
Libet, B. (1987). "Are the mental experiences of will and self-control significant for the performance of a voluntary act?" in *Behavioural and Brain Sciences*, 10
Libet, B. (1989). "The timing of a subjective experience" in *Behavioural and Brain Sciences*, 12
Libet, B. (1990). "Time delays in conscious processes" in *Behavioural and Brain Sciences*, 13
Libet, B. (1992). "Models of conscious timing and the experimental evidence" in *Behavioural and Brain Sciences,* 15:2
Libet, B. (2001). "Consciousness, Free Action, and the Brain. Commentary on John Searle's Article" in *Journal of Consciousness Studies*, 8 No. 8.
Locke, J. (1690). *An Essay Concerning Human Understanding*. Internet, but any edition will suffice
Lycan, W. G. (1992). "UnCartesian materialism and Lockean introspection" in *Behavioural and Brain Sciences,* 15:2
MacKay, D. M. (1985). "Do we 'control' our brains?" in *Behavioural and Brain Sciences,* 8:4
Marks, C. E. (1981). *Commissurotomy, Consciousness, and Unity of Mind*. Cambridge, Mass. MIT: Bradford Books
McCrone, J. "Libet replications and implications", *eScribe: PSYCHE-B: Internet*. Attached as an appendix.
McDermott, D. (1992). "Little 'me'" in *Behavioural and Brain Sciences,* 15:2
McGinn, C. (1979). "Action and its explanation" in **Bolton, N. (Ed.).** *Philosophical Problems in Psychology*. London: Methuen.
Merikle, P. M., and Cheesman, J. (1985). "Conscious and unconscious processes: Same or different?" in *Behavioural and Brain Sciences,* 8:4
Mortensen, C. (1985). "Conscious Decisions" in *Behavioural and Brain Sciences,* 8:4
Näätänen, R. (1985). "Brain physiology and the unconscious initiation of movements" in *Behavioural and Brain Sciences,* 8:4
Nagel, T. (1974). "What Is It Like to Be a Bat?", in **Block, N.** (1980). *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard.
Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford
Neethling, J., Potgieter, J. M., Visser, P. J. (1996). *Law of Delict*. Durban: Butterworths.
Nelson, R. J. (1985). "Libet's Dualism" in *Behavioural and Brain Sciences,* 8:4
Nietzsche, F. W. (1990). *Twilight of the Idols* and *The Antichrist*. London: Penguin
Nietzsche, F. W. (1990). *Beyond Good and Evil*. London: Penguin

O'Connor, T. (2002). "The Agent as Cause" in Kane, R. (Ed.) (2002). *Free-will*. Oxford: Blackwell.
Oxford. *The Concise Oxford Dictionary*. Oxford: Oxford.
Parfit, D. (1971). "Personal Identity" in *Philosophical Review,* Vol. 80
Parfit, D. (1984). *Reasons and Persons*. Clarendon: Oxford.
Pereboom, D. (2001). *Living Without free-will*. Cambridge.
Peterson, M., Hasker, W., Reichenbach, B., Basinger, D. (1996). *Philosophy of Religion, Selected Readings*. New York: Oxford
Plato. (380 BC). *Protagoras*. Internet.
Reid, T. In Lehrer, K. (1989), Honderich, T. (Ed.). *Thomas Reid*. London: Routledge
Rollman, G. B. (1985). "Sensory events with variable central latencies provide inaccurate clocks" in *Behavioural and Brain Sciences,* 8:4
Roskies, A. L., Wood, C. C. (1992). "Cinema 1-2-Many of the Mind" in *Behavioural and Brain Sciences,* 15:2
Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson
Sartre, J-P. (1992). *Existentialism and Humanism*. London: Methuen
Sdorow, L. M. (1993). *Psychology*. Dubuque: Brown and Benchmark.
Searle, J. R. (2001a). "Free-will as a Problem in Neurobiology" in *Philosophy: 76. 2001.*
Searle, J. R. (2001b). "Further Reply to Libet" in *Journal of Consciousness Studies:* 8, No. 8. 2001.
Searle, J. R. (1999). "Consciousness". Internet. *http://socrates.berkeley.edu/~jsearle/html/articles/consciousness.html.*
Skinner, B. F. (1948/2002). Extract from *Walden Two* in Kane, R. (Ed.) (2002). *Free-will.* Oxford: Blackwell.
Stamm, J. S. (1985). "The uncertainty principle in psychology" in *Behavioural and Brain Sciences,* 8:4
Stampe, D. W. (1992). "Of One's Own free-will" in *Philosophy and Phenomenological Research*, Vol LII, No. 3, September 1992.
Stich, S. P. (1990). *The Fragmentation of Reason*. Cambridge, Mass.: MIT
Strawson, G. (1994). "The Impossibility of Moral Responsibility" in *Philosophical Studies*, vol 75
Strawson, P. F. (1974). *Freedom and Resentment and other Essays*. London: Methuen.
Strawson, P. F. (1962). "Freedom and Resentment" in *Proceedings of the British Academy,* vol XLVIII.
Stump, E. (1993). "Intellect, Will, and the Principle of Alternative Possibilities" in Fischer, J. M., and Ravizza, M. (Ed.s) (1993), *Perspectives on Moral Philosophy*. Cornell University: Ithaca.
Teghtsoonian, R. (1992). *In defense of the pineal gland* in *Behavioural and Brain Sciences,* 15:2
Vanderwolf, C. H. (1985). "Nineteenth-century psychology and twentieth-century electro physiology do not mix" in *Behavioural and Brain Sciences,* 8:4
Van Gulick, R. (1992). "Time for more alternatives" in *Behavioural and Brain Sciences,* 15:2

Van Inwagen, P. (1983). *An Essay on free-will*. Oxford: Oxford.
Van Inwagen, P. (1989). "When is the will free?" in *Philosophical Perspectives*, vol 3.
Velleman, J. D. (2000). *The Possibility of Practical Reason*. Clarendon: Oxford
Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard
Wasserman, G. S. (1985). "Neural/mental chronometry and chronoethology" in *Behavioural and Brain Sciences,* 8:4
Watson, G. (1987). "Free Action and Free-will", in *Mind*, vol 96
Watson, G. (Ed). (1982). *Free-will*, Oxford: Oxford
Watson, G. (1975). "Free Agency" in *Journal of Philosophy*, vol. LXXII, No. 8, Apr.
Wolf, S. (1990). *Freedom Within Reason*. Oxford: Oxford
Wood, C. C. (1985). "Pardon, your dualism is showing" in *Behavioural and Brain Sciences,* 8:4