Self-Forming Actions, Snap Decisions, and Indeterminism:
A Problem for Kane's Libertarianism

Neil Campbell

Philosophy Dept., Wilfrid Laurier University

75 University Avenue W.

Waterloo ONN2L3C5, Canada

necampbell@wlu.ca

Abstract: Central to Robert Kane's libertarian account of free will is the distinction between two kinds of action: undetermined self-forming actions by means of which one shapes one's character, and those actions that are subsequently determined by one's freely formed character. Daniel Dennett challenges the coherence of this distinction, but I argue that his arguments rely on highly controversial assumptions. Building on one of Dennett's less problematic examples, I argue that many actions that appear to be determined by our character might very well turn out to be indeterministic. Since such actions occur in a way the agent seldom notices, yet meet the conditions Kane sets out for self-forming actions, they represent a problem for the degree of control agents have over their self-formation.

Robert Kane's (1998) brand of libertarianism has been extremely influential. Few theories have garnered as much attention or criticism in the contemporary debate about free will. For the most part, the criticisms of Kane's view fall into one of three groups. In the first, Kane's critics challenge the role indeterminism plays in his theory. They typically argue that indeterminism hinders rather than enhances the agent's control over her choices. This is because they believe indeterminism renders the outcome of deliberation random or arbitrary and frustrates the demand for contrastive explanations for why the agent didn't choose otherwise, despite

having the ability to do so (Mele 1999, 2005, 2006; Haji 2000; Levy 2005). The second group consists of those who challenge Kane's appeal to the dual efforts of will required for undetermined choices. Here the concern tends to be either that there is something incoherent about an agent trying to perform two incompatible actions at the same time (Clarke 1999; Allen 2005; Lemos 2011), or that dual efforts of will entail that the agent is responsible for too much, since her responsibility attaches to what she *tries to do*—not just to what she *actually does* (Levy 2005). The third group is made up of those who challenge the location of the indeterminism in Kane's theory. For reasons often associated with the first two groups of criticisms, these authors claim that the indeterminism arises too late in deliberation and argue that it should be installed earlier in the process in terms of non-actional states of the agent (Franklin 2013), the ordering of preferences (Ekstrom 2000, 2003), the weighting of considerations (Greenspan 2012), or acts of creative imagination (Pitman 2012), to name a few examples.

Although these criticisms are insightful and have pushed the discussion about libertarianism forward in helpful ways, there is an important dimension of Kane's theory that has been largely neglected. Few authors have critically examined Kane's basic distinction between two categories of action: the self-forming actions that require indeterminism, and those that are determined by the agent's freely formed character. This distinction is central to Kane's view since it allows for the possibility that undetermined choices are relatively rare, and hence, make up a small but significant group of actions. It is also fundamental to Kane's attempt to transform the free will debate from one about freedom of action into one about the possibility of self-formation. To my knowledge, only Daniel Dennett (2003) has pressed Kane on this distinction in any serious way. Like Dennett, I believe there is reason to think that the difference is not as well defined as Kane assumes. Unfortunately, Dennett's arguments rely on assumptions that are even more controversial than the distinction he attacks. In the light of this, I explore and develop a related but different line of argument in order to blur Kane's distinction. Drawing on one of Dennett's examples, I argue that some relatively innocuous assumptions about what appear to be central examples of non-self-forming actions ("snap decisions"), when coupled with the naturalistic framework of Kane's theory, imply that such choices are also undetermined. While this does not show that there is no difference between self-forming actions and actions that are determined by the agent's character, it does show that many more of our actions are undetermined than Kane suggests. More significantly, since undetermined snap decisions turn out to meet the conditions Kane sets out for self-forming actions, this introduces a problem for the kind of control agents can have over the project of shaping their characters. For now, it seems as though we can—perhaps to a quite significant extent—freely shape our characters in ways of which we are largely unaware. This is related to, but distinct from, the luck objec-

tion, since my central claim is not that the pervasiveness of undetermined choices exacerbates the problem of luck for Kane. Rather, the concern is that to the extent that such choices can happen tacitly and in the background, they normally escape our notice. This is an unwelcome result because this—rather than the indeterminism—undermines the control we have over such acts of self-formation.

My discussion is divided into three parts. In part I, I provide a brief sketch of Kane's theory and highlight his distinction between self-forming actions and those actions that are determined by one's freely formed character. In part II, I explore Dennett's attack on this distinction and explain why I find his criticisms problematic. In part III, I develop a related line of objection that is more promising and argue that it has important implications for Kane's views on the nature of self-formation.

1   Kane's Theory

Since Kane's theory is quite well known my stage setting will be relatively brief. Kane's is a variety of event-causal libertarianism that purports to be wholly naturalistic. Accordingly, he advocates only event causation and eschews any reference to agent causation or, indeed, any other "mysterious" forms of causation associated with noumenal selves or Cartesian mental substances. Kane's goal is to defend the idea that human beings can have ultimate responsibility (UR), whereby agents become the authors of their own characters, goals, and purposes. For this to be possible there must be room in the world for some degree of indeterminism within the realm of human choice. Otherwise, as Strawson (1994) claims in his 'basic argument', our characters would be determined by forces beyond our control. Kane calls the indeterministic choices by means of which we shape our own characters and thereby attain UR *self-forming actions* (SFAs)(Kane 1998, 74).

During an SFA the agent has a divided will, meaning that she has moral or prudential reasons that she endorses for two (or more) incompatible actions. Kane frequently describes this in terms of the agent being torn by two different visions of the kind of person she wants to be. Drawing on the idea that the brain employs parallel processing, Kane proposes that the competing reasons and the rationales that they support are physically realized in two corresponding neural networks that interact and compete for control of behaviour. In its bid for control, each network creates resistance to the opposing network, requiring the agent to make a dual effort of will to overcome both sources of resistance at once and to resolve the conflict. This battle for control between the two networks amplifies causal indeterminacies at the synaptic level, rendering the output of the interacting networks—the agent's choosing—undetermined, meaning that the agent could have chosen otherwise. Because the agent possesses what Kane calls "plural

voluntary control" (i.e., she endorses the reasons in favour of each choice and her choice is not compelled) the resulting action is unlike Austin-style examples where indeterminism undermines the agent's control (Kane 1998, 143). Whatever the outcome, the agent will have succeeded at doing something she was trying to do for reasons she accepts. Hence, according to Kane, it would be a mistake to think that the indeterminism involved in the process implies that the resulting action is involuntary or happens by accident.

The result of such SFAs is that they allow agents to set their wills in a certain way and thereby to secure responsibility for their own goals, purposes, or characters. When the businesswoman in Kane's favourite example resolves the conflict in her will between going to an important business meeting or stopping to help an assault victim, she shapes her character in such a way that disposes her to act a certain way in the future: selfishly or selflessly—depending on the outcome in this particular case.

Kane claims that UR is, or should be, the primary focus of the free will debate:

> "Focusing on UR helps to explain why I believe the tendency in the modern era to reduce the problem of freedom of the will to just a problem of free action is a mistake and oversimplifies the problem. Free will is not just about free action. It is about self-formation, about the formation of our wills (our characters, motives and purposes), or how we got to be the kinds of persons we are, with the wills we now have. Were we ultimately responsible to some degree for having the wills we do have, or can the sources of our wills be completely traced back to something over which we had no control, such as fate or the decrees of God, or heredity or environment, upbringing or social conditioning or hidden controllers, and so on? Therein, I believe, lies the core of the traditional problem of free will." (Kane 2014a, 40)

Once an agent is ultimately responsible, subsequent actions can be determined by the agent's freely formed character and flow from it effortlessly and without reflection. These later actions inherit their freedom from the provenance of the agent's previous SFAs.

> "It would follow [...] [from UR(N. Campbell)] that agents can act of their own free wills [...] even when their acts are determined and even when they could not have done otherwise—so long as the wills from which they act were formed by them by earlier SFAs that were not determined. Thus, incompatibilists can also say that acts done "of our own free wills" need not all be undetermined. Often we act freely and responsibly (even in an *incompatibilist* sense) out of a will already formed. What is required by UR is that we ourselves freely

formed that will, so that it is 'our own' free will." (Kane 1998, 78 *emphasis original*)

Returning to Kane's example of the businesswoman, if she decides to stop and help the assault victim, she shapes her will in such a way that her character *determines* subsequent actions. Although she could not do otherwise on these later occasions—given the kind of person she has become—these unreflective snap decisions are still free and responsible by virtue of being products of her freely formed disposition to help others in distress.

Kane's distinction between SFAs and those actions that are subsequently determined by the agent's freely formed will (henceforth, 'non-SFAs') is obviously central. In fact, this very distinction highlights the way in which he attempts to reshape the modern free will debate from one about freedom of action into one about self-formation. When we shape ourselves via SFAs we establish dispositions—ones of our choosing—that can operate more or less habitually. These habitual actions do not shape our characters, but flow from them. Hence, in Kane's view, our primary concern in the free will debate should really be on self-formation.

2   Dennett's Criticism

Although I will focus on Dennett's attack on the distinction between SFAs and non-SFAs, it is important to recognize that this is just one part of his more extensive critique of Kane's position. Dennett also argues against Kane's theory on two other fronts. He tries to show that Kane is guilty of treating the mind as a "Cartesian Theatre" by assuming that there is a place in the brain "where it all comes together"—a faculty of practical reasoning with clear boundaries between input and output where Kane can locate the indeterminism required for SFAs (Dennett 2003, 123). Such a view, according to Dennett, is fundamentally flawed because there is no such place (Ibid.). Dennett also challenges Kane's assumption that genuine indeterminism is required for SFAs by showing that deterministic pseudo-randomness would work just as well—if not better (Ibid.). I will not address these other lines of criticism here, but will focus exclusively on Dennett's attempt to collapse Kane's distinction between SFAs and non-SFAs.[1]

---

[1] I will, however, hazard two brief remarks about the other two prongs of Dennett's argument, though these remarks are not intended to be decisive. I don't find Dennett's claim that Kane's view assumes a Cartesian Theatre to be very compelling. Although Kane does make reference to a "self-network", and approvingly cites work by Crick and Koch (1990) about what might unify this network, it isn't entirely clear that such a picture is offensively Cartesian (Kane 1998, 137–142). After all, this network is distributed throughout the brain, and so are those networks involved in SFAs. Are these ideas any more implausible than neuroscientific theories that *localize* particular cognitive functions to specific regions of the brain? I find it difficult to see why they would be. As for Dennett's claim that pseudo-randomness will work just as well as genuine indeterminism, this might be susceptible to the concerns I identify about his verificationism later in section II, though I will not pursue this issue here.

Before launching into his argument, Dennett raises some initial doubts about this contrast. He employs the following example: His wife asks him if he can stop by the post office on his way to work to deliver a package for her. He says:

> "I reply almost instantaneously that I can't, because then I'd be late for an appointment with a student. Did I deliberate? Did I engage in a process of practical reasoning? This is not heavy-duty moral decision-making, but this is the stuff from which moral (and immoral) lives are largely composed: hundreds of thousands of minor choice points decided with a moment's consideration, usually with the background of justification kept tacit and unarticulated."

> (Ibid., 115)

Dennett then adds how odd it would be for him to reply to his wife as follows:

> "Well, since you are my wife and we have solemnly promised to help each other, and since I can think of no defect or problem in your request—you haven't asked me to do something physically impossible, or illegal, or self-destructive, for instance—there is undeniably a strong case for my answering 'Yes, dear.' On the other hand, I have told a student that I would meet with him at nine-thirty, and given the traffic, honouring your request would entail standing him up for at least half an hour. I could try to call him and ask his permission to reschedule, but I might not reach him, and besides, the harder question is whether my mailing the package in so timely a manner is a sufficiently important errand to warrant inconveniencing him. My making the appointment amounted to a promise to him, though not one that couldn't be forgivably broken for cause…". (Ibid.)

Dennett then makes the following interesting observation: "It is perhaps surprising to note that all these considerations (and many more!) really did contribute *somehow* to my snap answer" (Ibid.). He concludes that "[e]ven a snap judgement can be remarkably sensitive to myriad features of my world that have conspired over time to create my current dispositional state" (Ibid., 116).

The purpose of Dennett's example is to raise one of several boundary problems for Kane's theory. Here the boundary in question is the one between SFAs and non-SFAs. Dennett proposes that in the light of his alternative response to his wife it is unclear whether one should treat such snap decisions as issuing from the same faculty of deliberation as SFAs, or from a distinct faculty, leaving the former "in reserve for heavy lifting" (Ibid.). In a preliminary way, then, Dennett's example leads one to wonder just how an SFA differs from a non-SFA. Kane's answer, of course, is that only the former requires indeterminism. It is at this point in

his discussion that Dennett sets about undermining the contrast between these two types of action in earnest. His central argument is that there is no method or test by means of which we could tell the difference between an SFA and a non-SFA. Given this, the distinction is empty and ought to be abandoned.

At first blush, there appear to be two ways in which one might come to know whether or not a particular action is an SFA: either 'from the inside'—that is, from the perspective of the agent's own phenomenology of making a choice—or 'from the outside': using an empirical test or standard. According to Dennett, neither approach is helpful.

The problem with the first is fairly obvious and, given the framework of Kane's theory, quite compelling. Consider Dennett's Luther example. When Luther said, 'Here I stand. I can do no other', Kane agrees with Dennett that we need not treat Luther's inability to do otherwise as an obstacle to regarding his action as a free and responsible one. However, Kane disagrees with Dennett about the way in which this evaluation depends on Luther's earlier actions. Unlike Dennett, Kane thinks this judgement depends on the idea that Luther's action was determined by a will of his own making via earlier SFAs. We can certainly imagine Luther engaged in an inner struggle or two, becoming the kind of man he was when he finally uttered this statement. The difficulty is that from Luther's perspective it is a complete mystery whether or not his brain exhibited the required indeterminacy when he initially struggled with his allegiance to the Church of Rome to 'become his own man'. One cannot simply introspect and observe that two neural networks are interacting in such a way that they magnify indeterminacies at the synaptic level to produce an undetermined output. So the mere fact that one is *aware* of trying to choose—even in a tortured, hand-wringing way—is not a reliable indication of an SFA.[2] Hence, unless one can learn to introspect one's own brain states in the way that Paul Churchland (1985) suggests Frank Jackson's (1982) infamous Mary can, there is simply no way for an agent like Luther to know if he is in the throes of an SFA or not.

But even if one cannot tell from the inside that a particular choice is an SFA, surely one can tell from the outside, at least in principle. With the proper imaging technology, couldn't we detect an SFA since we have some idea of what we are looking for? Dennett claims we cannot: "when we get out our supermicroscopes and look at subatomic activity in the neurons, whatever we see will be equally uninformative about SFAs" (Dennett 2003, 129). Dennett justifies this bold assertion in two ways. The first hinges on the claim that genuine randomness is indistinguishable from pseudo-randomness; the second involves an example that illustrates our apparent unwillingness to allow whatever our 'supermicroscopes' tell us to override our ordinary practices of holding one another responsible.

---

[2]I will say more about the reliability of our introspective evidence below.

Pseudo-randomness has the appearance of genuine randomness but is fully deterministic. It is *practically* unpredictable—except for LaPlace's demon—due to the sheer number and complexity of the determining factors involved. Genuine randomness, on the other hand, is unpredictable *in principle*. Dennett's claim that what we observe in the brain couldn't possibly give us decisive evidence of an SFA is justified primarily by our inability to distinguish genuine from pseudo-randomness. Since we lack the cognitive abilities of LaPlace's demon, we simply can't tell if what we observe in the neural networks merely *appears* to be indeterministic, due to the sheer complexity of the neural connections, or *really is* indeterministic. Since we need to differentiate between these two possibilities in order to distinguish a non-SFA from a bona fide SFA, we can't do that either. Hence, even if our brain scans revealed the parallel processing Kane describes when an agent makes a difficult choice, and even if the output of these competing networks appeared to us to be undetermined, we might simply be wrong, so our 'supermicroscopes' are entirely unhelpful.

The other part of Dennett's justification for the claim that we can't identify an SFA using empirical means involves the following example. Suppose that a well-educated person with an exemplary upbringing is brought to trial for murder. Let us assume that throughout his life he has demonstrated normal moral development, understands the difference between right and wrong, and so possesses all the usual hallmarks of a responsible agent. Suppose that his team of lawyers 'demonstrates' that although he killed the victim, his brain is incapable of generating the indeterminacies Kane claims are required for SFAs. It would follow from Kane's theory that the defendant lacks UR, and so is not responsible for *any* of his actions, including the murder for which he is on trial. Dennett offers the following commentary on this example:

> "It's a tough sell. Why should the *metaphysical* feature of Ultimate Responsibility (supposing Kane has defined a coherent possibility) count more than the macroscopic features that can be defined independently of the issue of quantum indeterminism, and that are well motivated in terms of decision-making competences that agents have or lack? Indeed, why should metaphysical Ultimate Responsibility count for anything at all? If it can't be motivated as a grounds [sic] for treating people differently, why should anyone think it is a variety of free will worth wanting?" (Ibid., 131–132)

Since, according to Dennett, we would be loath to acquit the defendant on such metaphysical grounds—or better, to allow metaphysical considerations to trump our ordinary moral (and legal) standards of responsibility—even if we could detect the presence (or absence) of the indeterminism required for SFAs, our doing so would strike us as irrelevant to the issue of the defendant's responsibility. It follows that empirical evidence in the form of either the presence

or absence of indeterminism could not convince us of the difference between an SFA and any other action.

The above considerations lead Dennett to conclude that the distinction between an SFA and a non-SFA is specious, and so we should be highly sceptical of Kane's notion of ultimate responsibility. If there simply are no SFAs, they cannot serve as the source of our freedom of will or of action. In the light of this, Dennett thinks it is a non-starter to ground our responsibility in a handful of indeterministic choices. He proposes that if we want to retain something like the concept of an SFA *qua* an important character building choice-point in an agent's history, we should regard SFAs in the same way as speciation events. A speciation event, such as the occurrence of the first mammal, is identifiable only retrospectively and involves interest-sensitive properties of offspring as opposed to essential (in this case, mammalian) characteristics since there are only minute differences between each generation. *We* decide which differences are important, and our doing so necessarily involves a certain degree of arbitrariness. Similarly, Dennett proposes that an SFA is indistinguishable from other choices when they occur. Their significance becomes apparent only after the fact. The source of this significance is not that SFAs are *intrinsically* different from neighbouring choices, but is instead a matter of the way we organize such choices into the overall narrative of a person's moral life and development.

While I am sympathetic with the initial doubts Dennett raises about the contrast between SFAs and non-SFAs, I find his supporting arguments unconvincing. They are vulnerable on two fronts. The first front takes issue with Dennett's use of a verificationist principle; the second identifies some problematic assumptions at work in his example of the defendant whose brain lacks the capacity for indeterministic processes.

As we saw above, Dennett's main criticism of Kane's SFAs is that there is no way to detect them, and hence, no way reliably to differentiate between SFAs and non-SFAs. In the light of this, we should reject the distinction. Dennett is particularly fond of this style of argument. It relies on the following verificationist principle: *where there is no detectable difference, there is no difference*. Dennett has also employed this kind of argument to undermine the coherence of qualia (Dennett 1988) and of zombies (Dennett 1995). The trouble with this kind of approach, as William Seager has pointed out in the context of Dennett's discussion of qualia, is that the verificationist principle upon which it relies is at least as contentious as the concept under attack, if not more so (Seager 1999). This case seems to be no different. I think Kane could argue that we should reject Dennett's verificationist principle on the grounds that it is implausible to deny that there are distinctions we cannot reliably detect. To suggest otherwise exaggerates the relevance of our own limited abilities. By contrast, since there are plenty of theoretical reasons for thinking there are such things as SFAs—many of which are enumerated in his discussion

of parallel processing and the sciences of complexity and chaos (Kane 1998, 9–10, 117–118, 130, 186)—Kane can make a plausible case for the claim that we are *more* justified in believing in SFAs than we are in believing in Dennett's verificationist principle. Since verificationism is itself highly contentious, I see no reason to think that our inability to distinguish between genuine and pseudo-randomness should lead us to reject the distinction between SFAs and non-SFAs. Surely there can be such a difference in the absence of our ability to recognize it, and this is all that Kane's view requires. So, if we abandon Dennett's verificationist principle, it seems that much of his argument against Kane collapses.

One could reply on Dennett's behalf here that his argument does not really require the above verificationist principle, but just the claim that we are not in a position to know whether or not there are any SFAs, whereas Kane's view requires that we do know that there are such things as SFAs. While this is a tempting response, I don't think it succeeds. Kane is content to treat the existence of neural indeterminism (and hence, of the reality of SFAs) as an open question.[3] If determinism is true or if the only indeterministic processes that exist are not appropriately located in the human brain, Kane would concede that libertarianism is a lost cause. That determinism is true, or that the brain is a deterministic system are not settled issues. Until they are, it is reasonable for Kane to rely on the possibility of indeterministic neural processes without demonstrating their actuality. But then, in order for Dennett's initial criticism to hold, he must show that the concept of an SFA is incoherent, and his verificationist principle is designed to do just that.

The second problem with Dennett's argument holds even if one were to accept his verificationist principle. His example of the well-adjusted murder defendant is, in my view, unconvincing because it involves another problematic presupposition. His treatment of the example is driven by the assumption that metaphysical and neuroscientific considerations about free will and responsibility either do not or cannot inform our moral practices and judgements. Presumably, this is because he believes that such matters are entirely pragmatic, and hence, do not require any metaphysical moorings. If that sounds too strong, it is at least clear that Dennett expects most of us to agree that the defence's claim about the accused's lack of indeterministic neural processes is *irrelevant* to the assessment of his responsibility.

The trouble is that Dennett's expectation appears to be misplaced. In some intriguing recent work (Shariff, et al. 2014; Shariff & Vohs 2014) empirical research shows that when laypersons are taught about the free will debate and the neural basis of behaviour, their views on blame and responsibility change. In particular, findings suggest that such individuals prefer

---

[3]That this is the case comes across particularly clearly in his remarks (Kane 2014b) about Chapter 4 of Balaguer's (2010) recent book.

a less retributive form of punishment for criminals because they come to think that agents are less responsible for what they do than they originally thought. The conclusion of one such study offers the following, which has a clear bearing the way we should evaluate Dennett's example:

> "That mere exposure to modern neuroscience can be sufficient to reduce re-tributivist motivations may be particularly relevant to court cases. The explicit existence of free will may be rarely debated in court, but neuroscientific ev-idence often is. Indeed, recent research showed that judges afforded shorter sentences to hypothetical psychopathic criminals when the description of the criminals' psychopathy included a biomechanical component, compared with when it did not (Aspinwall, et al. 2012). Our findings likewise suggest that merely presenting such a perspective may move judges and jurors toward being less punitive and less retributive in general." (Shariff, et al. 2014, 1569)

In the light of this, it is perfectly reasonable to think that the jurors in Dennett's example would allow the fact that the defendant's brain is incapable of the indeterministic processes associated with SFAs to sway their decision towards a not guilty verdict. At the very least, this possibility should not strike us as incredible or absurd. Such evidence would not exist in a vacuum, after all. If we imagine (as Dennett's example requires us to) that neuroscience has advanced to the point where there could be conclusive evidence that someone's brain lacks the indeterministic processes required for SFAs, such advances would no doubt have begun to permeate our collective understanding. Given this, we can imagine the findings described by Shariff carried to the nth degree—of course the jury would find the defendant not guilty (Shariff et al. 2014). This suggests there is good reason to be deeply suspicious of Dennett's treatment of this example, and hence, of his support for the claim that empirical considerations about neural indeterminism are irrelevant to matters of responsibility, and so to the presence or absence of SFAs themselves.

These two considerations go a long way towards undermining Dennett's argument. If we reject Dennett's verificationism and treat the reality of SFAs as an open question, and if we disagree with Dennett's central example, his efforts to blur the distinction between SFAs and non-SFAs are not very persuasive. Since I am nevertheless sympathetic with Dennett's goal, I would like to explore another way of applying pressure to Kane's central distinction. My approach, however, will be quite different—almost the opposite, in fact. While Dennett tried to show that there are no such things as undetermined SFAs (and hence, that the distinction between SFAs and non-SFAs is specious) I will show that many of the actions that appear not to be SFAs in fact are. This is a considerably weaker conclusion than the one Dennett attempts

to motivate, but is no less important. For I will also argue that this conclusion has important implications for Kane's conception of self-formation.

3   SFAs, Snap Decisions, and Parallel Processing

We saw earlier that, according to Kane, once an agent has formed her character by means of SFAs, her character can determine subsequent choices such that they flow effortlessly from it. Following Dennett, let's call such choices 'snap decisions' since the idea is that the course of action in such cases is determined in an instant by the agent's character, and hence, bypasses the more complex indeterministic decision-making processes required by SFAs. My aim in this section is to show that even actions that appear to be snap decisions can have all the features of SFAs, and that this has troubling implications for Kane's views on self-formation. The argument, which is speculative (though no more so than Kane's own), will proceed in several steps. At each step, I will consider possible objections to the argument.

The first step is to show that, like SFAs, snap decisions actually involve deliberation and the consideration of competing reasons, all of which the agent endorses. Fortunately, we already have reason for thinking that this is the case in the light of Dennett's example of being asked by his wife to deliver a package. Upon closer examination, what initially appeared to be a non-deliberative snap decision turned out, in fact, to be much more complex. Many considerations, for and against, were actually at play, and presumably these all involved reasons Dennett endorsed. He believed that his duties as an academic and mentor provided compelling reasons to decline his wife's request, but he also believed that his duties as a partner and a husband provided a compelling reason to grant her request and to stand up his student. Hence, even in this simple case of a snap decision, Dennett displays the kind of plural rationality Kane associates with SFAs. It is also reasonably clear that this kind of thing is not an isolated phenomenon, for it seems fairly obvious that there is something to be said for and against virtually any action. Whether we notice it or not, there is reason to believe that there is a great deal of complex information processing that goes on in such cases, and hence, that before we act we are—on some level—deliberating between alternatives on the basis of conflicting reasons that we endorse.[4]

The Kanean is sure to object here and point out that the very fact one is unaware of anything resembling the kind of deliberation that occurs when one has a 'divided will' in cases like Dennett's snap decision only serves to highlight the difference between snap decisions and

---

[4]To be clear, the competing reasons are not simply in competition insofar as they recommend different courses of action. According to Kane, the reasons are "incommensurable" in the sense of being "noncomparable" (Kane 1998, 167). Hence, when I describe an agent as having competing reasons, I mean that these are normally incommensurable reasons.

SFAs. Dennett wasn't torn between two different visions of himself (the doting husband versus the dependable academic), nor did he have to make an effort of will in order to decide between delivering the package and meeting his student. His choice was effortless and instantaneous, which is surely very different from Kane's example of the businesswoman, who consciously struggles to resolve the conflict in her will.

There are two points here to address: the first involves the alleged differences between the speed at which SFAs and snap decisions occur; the second appeals to introspective or phenomenological differences between the two. Let's first deal with the *speed* at which Dennett makes his decision. The fact his choice was (or appeared to be) instantaneous is hardly decisive evidence that he did not deliberate about the alternatives or that his will was not divided. While it may be that some important choices are long drawn out affairs, the speed at which they occur is irrelevant to Kane's treatment of them as SFAs. Consider the businesswoman. She doesn't spend a week, a day, or even an hour deliberating about what to do—by then the crisis would be over. She has to decide then and there whether or not to intervene. Her choice, too, happens in an instant. Hence, the duration of the deliberation does not appear to be relevant to an action's status as an SFA—at least, not if we take Kane's favourite example at face value. So, the fact that Dennett's snap decision appears to happen in an instant does not show that it did not involve a divided will in just the way that SFAs do.

What, then, of the introspective differences between Dennett's choice and the businesswoman's? Surely the latter involves the subjective sense of a dilemma—even if it is just a momentary hesitation about what to do—while the former does not. Doesn't this show that Dennett's snap decision did not involve a divided will?

In response, I point to Kane's admission that introspection doesn't tell us the whole story about free will:

> "A frequently made objection is that we are not introspectively or consciously aware of making dual efforts and performing multiple cognitive tasks in self-forming choice situations. But I am not claiming that agents are introspectively aware of making dual efforts. What persons are introspectively aware of in SFA situations is that they are trying to decide about which of two options to choose and that either choice is a difficult one because there are resistant motives pulling them in different directions that will have to be overcome, whichever choice is made." (Kane 2014a, 44)

But if Kane is willing to allow that introspection doesn't tell us the full story about what is going on in an SFA, it is unclear why he insists that one must be introspectively aware of "trying to decide about which of two options to choose"(Ibid.). After all, introspection is

poorly understood and often unreliable (Nisbett & Wilson 1977), even in the context of making a choice (Nahmias, et al. 2004, Horgan & Timmons 2011). While the fallibility of introspection does not show that what it tells us about choice and deliberation can be completely discounted, the concerns about its reliability do raise serious doubts about how much introspection reveals about SFAs. After all, as we saw Dennett argue with the Luther example, even if one *is* aware of making a difficult choice, this does not guarantee the presence of the required indeterminism. Kane makes this concession himself when he discusses the similarities and differences between his theory and Balaguer's (2010) brand of libertarianism when he states the following about his own SFAs or Balaguer's torn decisions: "for all we know, decisions that feel torn might be determined" (Kane 2014b, 53). Similarly, one's lack of awareness of a deliberative conflict does not guarantee the *absence* of an indeterministic process. Hence, it does appear to be a genuine possibility that one can be in an SFA without being aware of this fact, and hence, without being aware of a deliberative conflict. Besides, if Kane claims that we are unaware of making dual efforts of will and are unaware of the cognitive processing that is going on in such situations, since the conflict *is* a matter of parallel cognitive processing it seems reasonable to expect that we can be unaware of the conflict itself. This allows for the possibility that Dennett's seemingly instantaneous snap decision really did involve a conflict of will, but one of which he was unaware—at least until he retrospectively thought about the situation more carefully.

We now have reasons to think that even Dennett's snap decision (1) involved a conflict in his will, and (2) included a process of deliberation that involved competing reasons that Dennett endorsed, and hence, plural rationality. The next stage of my argument requires the idea that two interacting neural networks physically realize these reasons. Here what I have to say will be highly speculative, but this seems fair given that Kane's own theory involves a largely hypothetical account of the supporting neurobiology. My suggestion may seem tongue-in-cheek, but is actually quite serious: *why wouldn't distinct but interacting neural networks realize Dennett's competing reasons?* Kane's reason for saying that the competing reasons in an SFA involve distinct networks is, as far as I can tell, because the reasons involve different mental contents, and hence, different information requires processing. The point of parallel processing, as Kane characterizes it, is to allow a complex system to deal with multiple kinds of information at once. Just as the colour and shape of an object are processed by different networks, so too are competing sets of reasons. Since a snap decision like Dennett's involves different reasons and contents, it seems reasonable to expect that two competing networks will also implement Dennett's conflicting reasons.

The Kanean will probably object that this part of Kane's theory involves a certain amount of conjecture, and so isn't it fair for him simply to *stipulate* that the way the competing reasons are physically implemented in an SFA differ from the way they are implemented in a snap decision? That is, can't he simply build the distinction into the underlying neurology and deny that snap decisions involve interacting parallel neural networks?

My response is, no, he can't. If we are to accept the claim that the way the competing reasons are physically realized is different for SFAs and snap decisions, this had better be because of some interesting and relevant differences between the cases and not a matter of sheer stipulation. Thus far, if my argument has been at all plausible, the differences between the businesswoman's SFA and Dennett's snap decision have not suggested *any* reasons for thinking that the way they are physically realized *must* differ in the way Kane seems to assume. We have yet to see any significant differences between the two cases that would suggest one involves parallel processing while the other doesn't. To claim otherwise is therefore entirely ad hoc.

We are now at the third and final stage of the argument. We need reasons to think that the competition between the two neural networks that realize Dennett's competing reasons magnifies quantum indeterminacies at the synaptic level, resulting in an undetermined choice. My tactic here is similar to the one in the previous step. If we agree that an SFA and Dennett's snap decision both involve parallel processing, *why wouldn't this result in an undetermined choice?* Presumably, the magnification of the synaptic indeterminacies is a product of the relations between the competing networks as they process their respective information and compete for global control. Since the information in a snap decision is implemented in the same way as that of an SFA (or so I have claimed), there is every reason to expect that it will result in an undetermined choice too. To suggest otherwise is, once again, ad hoc. Given Kane's conjecture about the way in which conflicts of will disrupt thermodynamic equilibrium in the brain, there is every reason to suppose that this happens during snap decisions as well. After all, either the resulting indeterminism is a product of the physical system, or it isn't. If it is, and we have reason to think the system is structurally the same for SFAs as it is for snap decisions, then the latter will also result in undetermined choices; if it isn't, then the indeterminism is the product of something else that looks suspiciously like magic, and Kane can no longer claim that his theory is naturalistic.

This completes the argument. By combining Dennett's example with the same naturalistic assumptions that drive Kane's theory, I have shown that there is good reason to think that even snap decisions have the same features as undetermined SFAs. If we assume that the kind of phenomenon that Dennett describes is not an isolated one, this suggests that far more of our choices are undetermined than Kane proposes. This is a more modest conclusion than

Dennett's, for it doesn't collapse Kane's distinction between SFAs and those non-SFAs that are determined by one's freely formed character. To do that one would have to show that *all* non-SFAs have the same features as the snap decision described above. I will not attempt this.

So how serious of a problem is this weaker conclusion for Kane's theory? Since he is a libertarian and embraces the idea that some of our free choices are undetermined, couldn't he welcome this conclusion? Wouldn't *more* indeterminism be *better*? I don't think so—at least, not if he is serious about reconceiving the free will debate as one about the possibility of self-formation. The problem I see for Kane is this: if snap decisions turn out to have all the hallmarks of SFAs, then they have an equal claim to a role in self-formation. This is problematic because, as we saw, snap decisions involve deliberations that become apparent to the agent only retrospectively, if at all. At the time of their occurrence, the agent is unaware of making such choices. This is a peculiar result, for it means that we make free choices and shape our characters in ways that do not involve conscious or reflective deliberation and control. Although such choices serve the same regress-stopping function as SFAs, they appear to diminish rather than to enhance the amount of control we have over self-formation.[5]

This is quite problematic for Kane's view, for it is similar to the problem he identifies for Frankfurt's approach to free will (Frankfurt 1972, 1987). Kane describes his difference of opinion with Frankfurt as follows:

> "Many years ago, I wrote a short letter to Frankfurt which mentioned several criticisms of his view. The familiar criticism was that whether or not we were wholehearted about the will we had, rather than ambivalent, could on his view be entirely a matter of social conditioning over which we had no control. The second, less familiar criticism was this: if freedom of will is being wholehearted in your commitments to what you will, without ambivalence, then no one could ever get from ambivalence to wholeheartedness 'of their own free will'. For they wouldn't *have* free will on his view until they got there." (Kane 2014a, 38)

Frankfurt answered forthrightly by saying he believed it didn't matter how you came to have free will or wholeheartedness. It might only be because of good fortune or upbringing or accidental factors or even social conditioning. He adds:

> "And no, you can't get to free will from ambivalence of your own free will in any deeper sense of that term. All that matters is that you have free will

---

[5]Hence, the nature of my concern isn't one about the sheer number of SFAs or about the pervasiveness of undetermined choices, though to the extent that critics like Mele are correct that indeterminism undermines control, the ubiquity of undetermined choices would be a problem. Setting aside the effectiveness of the luck objection, my worry is that if we engage in self-forming actions about which we are unaware, it will turn out that we are freely shaping our characters in ways that lack conscious control, and this is a peculiar result.

or wholeheartedness no matter how you got it, for it is a good which makes life go well. This honest answer did nicely frame the debate between us, as compatibilist and incompatibilist, about what acting of one's own free will might mean." (Kane 2014a, 38)

Kane, then, thinks it is a *failing* if a theory about self-formation (or having the will one wants) allows for the possibility that how one comes to be that self (or have that will) is not under one's own control. Yet this is precisely what the above argument about snap decisions suggests since the agent engages in a form of self-formation that passes largely unnoticed. Although snap decisions are not the product of Frankfurt-style controllers or social conditioning—they are undetermined, after all—the fact we are largely unaware of them surely suggests they are not under our control in the way Kane thinks most SFAs are. This conclusion threatens Kane's account of self-formation, and so he needs to address the above possibility.

As I see it, Kane has two options: he can either undermine the significance of the above argument by marginalizing the effect that undetermined snap decisions have on our control over the process of self-formation, or he can show that there are principled reasons to think that snap decisions are not undetermined and lack the general features of SFAs. In the light of Kane's reaction to Frankfurt it is difficult to see how he could adopt the first approach. It is central to his view of self-formation that we have a certain amount of control over such acts. I suspect the second is more promising, but unless Kane can show that introspection is sufficiently reliable to distinguish SFAs from snap decisions, or can muster neurological evidence to show that snap decisions are (or must be) physically realized in a different way than SFAs, there is little reason for optimism about this second approach.

In conclusion, I have argued that Kane's libertarian theory faces a puzzling problem. Drawing on Dennett's attempt to undermine Kane's distinction between SFAs and non-SFAs, I have shown that there is reason to suspect that many snap decisions will turn out to have all the same features as SFAs. Both involve a conflict of will between reasons the agent endorses, both involve parallel processing of these respective reasons, and unless Kane surrenders his naturalism, it seems that the outcome of both processes is undetermined. While this might initially appear to benefit Kane's brand of libertarianism, I have suggested that it actually has the opposite result since such snap decisions undermine the control agents have over the formation of their own characters. In the light of Kane's reaction to Frankfurt's views on self-formation, this is surely a result he cannot accept.[6]

---

[6]I wish to express my gratitude to two anonymous referees for their helpful and insightful comments on earlier drafts of this paper.

References

Allen, R. (2005), 'Free Will and Indeterminism: Robert Kane's Libertarianism', *Journal of Philosophical Research*, **30**, 341–355.

Aspinwall, L. G., Brown, T.R. & Tabery, J., (2012), 'The Double-Edged Sword: Does Biomechanism Increase or Decrease Judges' Sentencing of Psychopaths?', *Science*, **337**, 846–849.

Balaguer, M. (2010), *Free Will as an Open Scientific Problem*, Cambridge, MA: MIT Press.

Churchland, P. (1985), 'Reduction, Qualia, and the Direct Introspection of Brain States', *Journal of Philosophy*, **82**, 8–28.

Clarke, R. (1999), 'Free Choice, Effort, and Wanting More', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **2**, 20–41.

Crick, F., & Koch, C. (1990), 'Towards a Neurobiological Theory of Consciousness', *Seminars in the neurosciences*, **4**, 263–276.

Dennett, D. C. (1988), Quining Qualia, *In* Marcel, A.J., & Bisiac, E. (eds.), 'Consciousness in Contemporary Science', Oxford: Clarendon, 381–414.

———. (1995), 'The Unimagined Preposterousness of Zombies', *Journal of Consciousness Studies*, **2**, 322–325.

———. (2003), *Freedom Evolves*, New York: Viking.

Ekstrom, L. W. (2000), *Free Will: A Philosophical Study*, Boulder, Colo.: Westview Press.

———. (2003), 'Free Will, Chance, and Mystery', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, **113**, 153–180.

Frankfurt, H. (1972), 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, **68**, 5-20.

———. (1987), Identification and Wholeheartedness, *In* Shoeman, F. (ed.) 'Responsibility, Character, and the Emotions', Cambridge: Cambridge University Press, 27–45.

Franklin, C. E. (2013), 'How Should Libertarians Conceive of the Location and Role of Indeterminism?', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **16**, 44–58.

Greenspan, P. (2012), 'Free Will and Rational Coherency', *Nous-Supplement: Philosophical Issues*, **22**, 185–200.

Haji, I. (2000), 'Indeterminism, Explanation, and Luck', *Journal of Ethics: An International Philosophical Review*, **4**, 211–235.

Horgan, T., & Timmons, M. (2011), 'Introspection and the Phenomenology of Free Will: Problems and Prospects', *Journal of Consciousness Studies*, **18**, 180–205.

Jackson, F. (1982), 'Epiphenomenal Qualia', *Philosophical Quarterly*, **32**, 127–136.

Kane, R. (1998), *The Significance of Free Will*, New York: Oxford University Press.

———. (2014a), 'Acting 'of One's Own Free Will': Modern Reflections on an Ancient Philosophical Problem', *Proceedings of the Aristotelian Society*, **114**, 35–55.

———. (2014b), 'Torn Decisions, Luck, and Libertarian Free Will: Comments on Balaguer's Free Will as an Open Scientific Problem', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, **169**, 51–58.

Lemos, J. (2011), 'Wanting, Willing, Trying, and Kane's Theory of Free Will', *Dialectica: International Journal of Philosophy of Knowledge*, **65**, 31–48.

Levy, N. (2005), 'Contrastive Explanations: A Dilemma for Libertarians', *Dialectica: International Journal of Philosophy of Knowledge*, **59**, 51–61.

Mele, A. R. (1999), 'Kane, Luck, and the Significance of Free Will', *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, **2**, 96–104.

———. (2005), 'Libertarianism, Luck, and Control', *Pacific Philosophical Quarterly*, **86**, 381–407.

———. (2006), *Free Will and Luck*, Oxford: Oxford University Press.

Nahmias, E., Morris, S., & Nadelhoffer, T. (2004), 'The Phenomenology of Free Will', *Journal of Consciousness Studies*, **11**, 162–179.

Nisbett, R. E., & Wilson, T. D. (1977), 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review*, **84**, 231–259.

Pitman, M. M. (2012), 'Freedom, Indeterminism and Imagination', *South African Journal of Philosophy*, **31**, 369–383.

Seager, W. (1999), *Theories of Consciousness: An Introduction and Assessment*, London: Routledge.

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B.,Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (2014), 'Free Will and Punishment', *Psychological Science*, **25**, 1563–1570.

Shariff, A. F., & Vohs, K.D, (2014), 'The World without Free Will', *Scientific American*, **310**, 76–79.

Strawson, G. (1994), 'The Impossibility of Moral Responsibility', *Philosophical Studies*, **75**, 5–24.