

“The apparent illusion of conscious deciding”

Joshua Shepherd

ABSTRACT

Recent work in cognitive science suggests that conscious thought plays a much less central role in the production of human behavior than most think. Partially on the basis of this work, Peter Carruthers has advanced the claim that humans never consciously decide to act. This claim is of independent interest for action theory, and its potential truth poses a problem for theories of free will and autonomy, which often take our capacity to consciously decide to be of central importance. In this paper, I examine the nature of conscious deciding and I argue that Carruthers fails to establish the claim that humans never consciously decide to act.

1. Introduction

Recent work in cognitive science suggests that conscious thought plays a much less central role in the production of human behavior than most think (Wilson, 2002; Wegner, 2002). Partially on the basis of this work, Peter Carruthers – in an impressive series of recent papers (2007, 2009a, 2009c, 2009d, 2009e, 2010) – has advanced the claim that humans never consciously decide to act. This claim is of independent interest for action theory, and its potential truth poses a problem for theories of free will and autonomy, which often take our capacity to

The author would like to thank Alfred Mele, Steve MacFarlane, Zachary Martin, Brendan Ritchie, and two anonymous reviews as well as the audience at the 2011 meeting of the Society for Philosophy and Psychology for helpful feedback on earlier versions of this manuscript.

consciously decide to be of central importance (O'Connor, 2009; Mele, 2010). In this paper, I outline Carruthers' arguments, and I argue that they fail to establish the claim that humans never consciously decide to act.

I should note that I do not offer an argument that we ever do consciously decide to act. For all I say below, Carruthers' claim could turn out to be true. My position is that Carruthers' arguments do not establish the claim.

2. Deciding

It will help to briefly discuss the nature of decisions to act, or *practical* decisions. These decisions are about what to do – i.e., whether to do something, when to do something, or how to do something. Practical decisions are distinct (even if related in important ways) from decisions about what is the case. Most philosophers who have written about practical decisions take them to be mental actions of forming intentions (Clarke, 2003, pp. 3-27; Mele, 2003, pp. 197-202; Frankfurt, 1988, pp. 174-176; Kane, 1996, p. 24; McCann, 1986, pp. 254-55; Searle, 2001, p. 94). Some take practical decisions not to be actions, but rather to be events of non-actional intention acquisition (O'Shaugnessy, 1980, pp. 300-301). But Carruthers seems to side with majority opinion, maintaining that decisions are “acts¹ of willing, or the events that create novel activated intentions” (2010, p. 78).² In what follows, then, I will adopt with Carruthers the majority view that decisions are mental actions of intention formation, and I will consider his case for the claim that such actions are never performed consciously.

¹ Some action theorists use the term ‘mental act’ as a noun, using it to refer to mental phenomena many would regard as non-actional, such as seeing and hearing (Geach 1957). At issue here is a conception of deciding on which a decision is an action an agent performs. To avoid transgressing a distinction between acts and actions, I speak in what follows of the mental action of deciding.

² Notice that here Carruthers treats actions as events. In doing so Carruthers sides with a majority of philosophers of action, although some argue that actions are not events (e.g., Bach 1980). Since conceiving of actions as events gives Carruthers' arguments the most purchase for the philosophy of action, I will follow Carruthers in doing so.

Regarding conscious decisions, two issues need clarification. First, *how* does an agent consciously form an intention? Since Mele's (2009) account of conscious practical deciding would be falsified by the truth of Carruthers' claim, it will be useful to assume it for dialectical purposes in what follows. On this account, when an agent decides to act, an agent forms an intention by performing a basic conscious mental action of assent to an action-plan, where the action of assent is not to be reduced to some other mental operation, such as uttering 'I will do X' in inner speech (Mele, 2009, pp. 40-44, cf. Mele, 2010).

Second, what does it mean to assert that a decision is *consciously* performed? Full consideration of issues relevant to this question would take us far afield. Fortunately, Mele (2009) offers an explication which is useful for present purposes. On Mele's explication, an agent consciously decides when she is aware of performing the action of deciding the whole (of course brief) time she is performing it (2009, p. 43). Importantly, the agent is not aware of her decision in the way she is aware of an external event – for external events begin some brief time before we become aware of them. Rather, the agent is aware of the onset of the action from the very onset. As Mele puts it:

Deciding to A . . . is not to be confused with any process that issues in a decision to A, including, for example, deliberation about what to do. And deciding to A does not precede the onset of the intention to A formed in the act of deciding. Instead, what it is to decide to A is to form – actively – an intention to A. The intention arises *in* that momentary intention-forming action, not after it. (2009, p. 43)

In what follows, I examine the claim that a certain kind of mental action cannot be consciously performed. This is a restricted question, but it retains interest. The capacity to perform such actions is plausibly related to other conscious mental actions, such as imagining, attempting to remember, and directing attention. The capacity to consciously decide is arguably basic to any view on which the conscious mind is active, rather than a mere witness of mental activity.

3. The Propositional Attitude Argument

Carruthers' position on decisions can be seen as one consequence of his more general position on the nature and origin of self-knowledge. For Carruthers, we come to know our own thoughts and actions by way of what he calls our 'mindreading system' – a suite of cognitive mechanisms the function of which is to monitor various internal and external events and use the incoming evidence to attribute mental states. Carruthers argues that the human mindreading system is a single purpose system, discerning the mental states of the mind in which it operates in the same way as it discerns the mental states of other minds. According to Carruthers, "When we attribute mental states to ourselves we utilize the same conceptual and inferential resources that we use when interpreting others, with the result that our only access to a significant class of our own mental states is self-interpretive rather than introspective" (2009e, p. 468).

Carruthers grants that when we attribute mental states to ourselves, our mindreading system benefits from having direct access to *more information* than it does in the case of others. For example, we can introspect parts of the mental images rehearsed in consciousness, as well as our own inner speech and various somatosensory states generated by emotions. What we lack is direct access to a certain 'significant class' of mental states – namely, our propositional attitudes.

What I call the Propositional Attitude Argument applies the above view on self-knowledge to decisions. We can represent the argument as follows.

1 If we lack introspective access to a mental event, then that mental event is not a conscious event.³

2 We lack introspective access to propositional attitude events.

3 A decision to act is a propositional attitude event.

C Thus, no decision to act is a conscious event.

In order to assess this argument, it is important to understand the sense in which decisions might be propositional attitude events. This question is bound up with a further question, namely, what is a propositional attitude event? Carruthers says little about this, but this seems a fair interpretation: a propositional attitude event is the event of an agent's forming or acquiring a propositional attitude, for example a belief, desire or intention. Assuming this is correct, in what sense is a decision a propositional attitude event?

Two undesirable readings of 'propositional attitude event' suggest themselves. First, one might think that decisions *are* propositional attitudes, and thus that deciding is the event of acquiring or forming a decision. The locution 'decision' is sometimes used in this way, to speak not of the *action* of deciding, but rather of the *content* of the decision. But this reading will not do. For decisions are actions, and actions are not propositional attitudes. Second, one might think

³ As Carruthers acknowledges (2010, p. 77), those who endorse a first-order view of consciousness – according to which phenomenal consciousness is constituted by those contents which are available to first-order processes for the guidance of thought and action (see, e.g., Tye 1995) – would deny this claim. For on a first-order view, it is possible that mental events which are not introspectable could nevertheless be conscious. Here I bracket concerns about premise 1, and focus on premises 2 and 3. Concerns relevant to Carruthers' view of consciousness reappear in the next section.

that decisions are the events of acquiring or forming a propositional attitude *related* to the decision. But this option is too broad. For we often form beliefs related to our decisions – e.g., the belief that ‘I have decided to Q.’ Since decisions are actions of intention formation, the event of deciding is distinct from the event of forming a belief. We might conjoin them, of course, but the resulting conjunction is *ad hoc*. Notice that on this reading, Carruthers’ argument will not go through. For it will be left open that even if we have no immediate awareness of belief-formations, the same is not true of decisions.

A third reading is available, however. On this reading, a decision is a propositional attitude event because intentions are propositional attitudes, and a decision is the formation of an intention. This seems to be the most plausible reading. For when philosophers of psychology talk about propositional attitudes, they sometimes put intentions on the list. Are they right to do so? Our answer to this question will depend on how we wish to analyze propositional attitudes – a contentious issue I would rather avoid here (but see Schroeder, 2006). In what follows, I will assume that the intentions we are talking about are propositional attitudes.

Consider the argument’s second premise. Carruthers’ case for it is abductive, drawing on a wide range of evidence from cognitive science. The case is both thorough and impressive, but here I need not worry about it in full. The question is whether the more general claim about awareness of propositional attitude events covers a particular kind of propositional attitude event, namely, a decision. It is possible that Carruthers is right about certain kinds of propositional attitude events (e.g., judgments) and not others.

The evidence most directly relevant to the claim at issue stems from the literature on confabulation. Carruthers argues, in general, that confabulation about propositional attitude events indicates that such events are not introspectively accessible. But it is worth noting that our

question considerably restricts the relevant data. We want to know whether, why and how often people confabulate about *decisions*. Much of the literature on confabulation focuses on confabulation of reasons for or causes of *overt* action or confabulation of reasons for or causes of *judgment*. But evidence for confabulation about the causes of or reasons for overt action or judgment does not necessarily transmit to the case of practical deciding.

Consider, for example, a classic study by Nisbett and Wilson (1977). In the guise of a consumer survey, Nisbett and Wilson had participants observe four pair of nylon stockings, situated from left to right in front of them. The participants were asked to judge which pair “was the best quality” (1977, p. 243). Although the stockings were identical in all relevant respects, participants chose the pair on the right by a factor of almost four to one. When asked to explain their judgment, participants never mentioned the position of the stocking as an influence on their judgment. This is some evidence that participants were not fully aware of the processes that led to their judgment, and thus that confabulation occurred. But this does not indicate that participants were unaware of *any* reasons that led to their judgment, a point Carruthers acknowledges (2010, p. 95). Consider a father (Dan) deliberating about whether to give the last ice cream sandwich to his son or his daughter. Those who know Dan might predict that the son has no chance – he is biased in his daughter’s favor. Suppose that Dan tells himself he will give the ice cream sandwich to the most deserving child, and he deliberates for a second about this. Suppose as well that Dan’s decision is made primarily because he visualizes his daughter’s joy upon receiving the ice cream sandwich. In this case Dan is not aware that his bias drives his decision, but he is rudimentarily aware of the reason that he chose his daughter.⁴ He did so

⁴ It seems possible that Dan could be aware of no reason to give the ice cream sandwich to his daughter over his son, and yet be aware of his performing the action of deciding. Indeed, such a possibility might be actual in neuroscientific studies which involve participants moving a finger at will (e.g., Brasil-Neto et al. 1992). Awareness

because his visualization of her joy made him happy. Dan might confabulate as he explains to his son or wife how much his daughter deserved the ice cream sandwich. But such confabulation does not indicate that he did not consciously decide to give the ice cream sandwich to his daughter.

This is an instance of a more general point regarding an agent's awareness of her actions.⁵ As work on the role of vision in overt action has made clear, we are often unaware of the fine-tuned adjustments we make when performing overt actions – e.g., when reaching for a mug of coffee (see Milner and Goodale, 2006). This does not indicate that we are unaware of our actions under *some* description. I might be aware of my action of reaching for a mug of coffee, even though I lack awareness of the programming of that action's kinematics, as well as the unconscious visual states being used to adjust fine-grained elements of that action. In other words, I might be aware of what I am doing under some descriptions, but not under others.

Something similar is arguably true of decisions. In cases of confabulation the agent is unaware of *what* was decided under some descriptions. For example, we might say that Dan decided to give the ice cream sandwich to his favorite child. And Dan might say he decided to give it to the most deserving child. Neither description makes sense of what was *consciously* decided. This does not indicate that nothing was consciously decided, however. In the above story Dan consciously decided (a) to give the ice cream sandwich to his daughter, (b) because the conscious thought of doing so made him happy. Though Dan was unaware of what was decided under a number of descriptions, he was aware of what was decided in this more restricted sense.

Regarding practical deciding, then, we want to know whether instances of confabulation warrant the inferences that (a) the agents in question did not consciously decide anything, though

of reasons for deciding should be kept distinct from awareness of deciding. Thanks to XXXX for bringing my attention to this point.

⁵ Thanks to an anonymous reviewer for suggesting this connection.

they claim they did, and (b) whether such confabulation plausibly generalizes to all instances of conscious deciding. Carruthers claims that there is “ample evidence” of confabulation for decisions, and cites two studies in support of this claim (2009c, p. 131; cf. also 2010, pp. 88-92). Each study merits brief discussion.

In the first study (Brasil-Neto et al. 1992), participants were instructed to decide whether to move their right or left index finger after hearing a click. The click emanated from a Caldwell MES-10 magnetic stimulator, which in one condition was directed away from participants, and in another applied transcranial magnetic stimulation (TMS) to prefrontal or motor areas of the brain. When stimulation was applied to motor areas of the brain, 80 percent of participants who moved their finger within 200 ms of the click chose the finger contralateral to the site stimulated. Brasil-Neto et al. report that these participants were not aware of any influence of TMS on their finger choice, but they do not give any further information – it is thus unclear how participants described their lack of awareness. Participants who moved their finger 200 ms or longer after the click (some took longer than a second to decide) did not show bias for one finger or another. In other words, the relevant effect involved only participants who responded in under 200 ms.

Brasil-Neto et al. speculate as follows. “In the trials in which the magnetic stimulus had an effect, it is likely that activation of the response channel during the foreperiod came very close to the threshold, with the result that the magnetic stimulus aided the production of the response” (1992, p. 965). Given the likely causal role of the TMS, coupled with the lack of awareness, Brasil-Neto et al. suggest that “it is possible to influence movement preparation processes externally without disrupting the conscious perception of volition . . . [our results] suggest that conscious perception of willing a particular action is preceded, and possibly generated, by cerebral processes that can be influenced by magnetic stimulation” (1992, p. 966).

In his interpretation of this study Carruthers claims that Brasil-Neto et al. “caused subjects to move one index finger or another . . . Yet the subjects themselves reported *deciding* to move the finger” (2009c, p. 131). The implication is apparently that participants’ reports are confabulatory since TMS applied to motor areas, as opposed to any conscious activity, is responsible for the finger movement. This is misleading. First, Brasil-Neto et al. offer evidence that TMS played a role – i.e., by raising the proportion of contralateral finger movements to 80 percent – in the finger movements of *some* participants, namely those that moved their fingers shortly after the click. This is far from the full causal manipulation implied by Carruthers. Second, it is unclear *what* the subjects reported, since Brasil-Neto et al. do not say. Even if subjects thought they were consciously deciding to move the finger they moved, it is unclear why we should take this thought as illusory. Plausibly, all decisions are influenced by prior nonconscious events. Is there any reason to think that this alone renders the experience of such decisions illusory?

Carruthers thinks so. He argues “everything we know about the organization of the brain suggests that motor cortex isn’t the place where decisions are themselves made, nor is there any plausible route via which stimulating motor cortex could cause a decision to be made” (2010, p. 88). The idea is that TMS to motor cortex caused the finger movement, with no causal input from decision-making systems. But, again, stimulating motor cortex did not cause a finger movement in all cases – rather such stimulation significantly raised the proportion of contralateral finger-raises in participants who responded within 200 ms of the stimulation. Given this, Brasil-Neto et al.’s interpretation suggests an explanation consistent with the hypothesis that agents sometimes consciously decide. Recall the interpretation: participants who moved their contralateral finger within 200 ms, the activation of the response channel was already very close

to the threshold for that finger, such that the TMS pushed these participants over the top. Now, one might argue that since the agent's decision-making system is anatomically distinct from the response channel in motor cortex, the activation of the response channel has little to do with a participant's decision-making process. But in fact a recent study by Selen et al. (2012) indicates otherwise.

In this study Selen et al. demonstrate that motor cortex is in close communication with decision-making systems throughout the deliberational process, such that “there is continuous flow from the decision making process to set feedback gains in the motor system” (2012, p. 2284). In the experiments participants had to judge the general direction of motion (left or right) for a group of randomly moving dots. Participants indicated their decision by moving a joystick left or right. Selen et al. perturbed participants' arms at various times during the task, and measured the strength of their reflexes by attaching electrodes to participants' arms. They found that the more information subjects received – and thus, the closer participants were to deciding, as measured by a decision variable – the more intense their reflexes upon perturbation. As Selen et al. explain, thanks to the continuous flow of information from decision making systems to motor cortex, “there is supraspinal modulation of [reflex] gain during decision making, reflecting changes in the excitability of the response pathways” (2012, p. 2284).

Applying this new result to Brasil-Neto et al.'s renders plausible the following possibility. In participants who acted quickly (that is, within 200 ms of the stimulation), increased activation in the response channel indicated that participants were already nearing a decision to move the contralateral finger. This possibility would explain why such agents would not experience the TMS as interfering with their decision-making process. Further, this explanation is consistent with the possibility that participants who did not respond within 200 ms

consciously decided to move their finger. Ultimately, Brasil-Neto et al.'s study offers little support for premise 2, as applied to decisions.

In the second study Carruthers cites, Wegner and Wheatley (1999) had subjects collaborate with a confederate by sitting opposite them at a table, and by controlling one half of a wooden square which sat atop a computer mouse. By moving the square, the subject and confederate moved the mouse – and as they did so they observed a cursor moving across a computer screen populated by about 50 objects (e.g., a swan, a car, and so on). The subjects were instructed to stop the mouse on an object every 30 seconds or so. In addition, the subject and confederate wore headphones which played music and words. The music indicated that the subjects should prepare to stop the board, and the words were presented ostensibly as random distracters. In reality, the words served to prime objects on the screen (e.g., 'swan,' 'car').

In some cases – unbeknownst to the subject – the confederate stopped the square (forced stop cases). In others, the subject was allowed to stop the square (unforced stop cases). Then subjects were asked to report their degree of intentionality where 0 indicated 'I allowed the stop to happen' and 100 indicated 'I intended to make the stop.' On unforced stop cases, subjects reported about 56% intentionality. On forced stop cases, subjects' degree of intentionality varied with the timing of the prime word. When the prime happened 30 seconds before the stop, subjects reported about 44% intentionality. When the prime happened 5 seconds before the stop, intentionality was rated at about 60%. Thus the priming had a significant effect of the subjects' report of intentionality. Wegner and Wheatley conclude that "the experience of will can be created by manipulation of thought and action . . . and this experience can occur even when the person's thought cannot have created the action" (1999, p. 489). Carruthers interprets the experiment as follows: "It seems that the subject's mindreading faculty, presented with the

evidence that the subject had been thinking of the relevant object shortly before the cursor came to a stop beside it, reasoned to the most likely explanation, and concluded that the subject had taken a decision to stop beside that very object” (2009c, p. 131).

Are these interpretations warranted? And do they undermine the trustworthiness of our experience in cases of (seeming) conscious decisions to act? There are reasons to think not. First, the subject’s experience of intentionality was rather low. Even in unforced stop cases, the rating of 56% indicates a degree of confusion about what, if anything, was decided. Second, it is not clear that the subjects attributed a *decision* to themselves. We have no data on how subjects interpreted the scale on which their answers were recorded. Perhaps they simply felt that the action occurred thanks in part to their participation. One expects that if subjects were allowed to stop the board themselves, after deciding to do so, ratings would approach 100%. So it seems we’re owed an explanation for the difference in confidence.⁶

It is of course true that the confabulation literature suggests that we often lack full awareness of the factors which give rise to decisions. But this is very different from the claim that we never consciously decide. It may turn out that new studies indicate much more pervasive confabulation regarding decision-making. But short of the existence of such studies, the confabulation literature does little to advance the claim that we never consciously decide.

4. The Causal Role Argument

⁶ Indeed, such a point applies to Carruthers’ use of the confabulation literature more generally. As Fiala and Nichols note, most confabulations are reported with what seems to be low confidence. Often, participants will respond to a question by answering ‘I don’t know,’ before beginning to confabulate. Fiala and Nichols argue that this seeming difference in confidence offers “prima facie reason to think there are systematic differences in confidence levels between confabulation and apparent introspection, which in turn suggests a difference in underlying mechanism” (2009, p. 25).

I have argued that the confabulation literature does little to advance the claim that we never consciously decide. Carruthers' case for this claim does not rely on the confabulation literature alone, however. The argument under consideration in the next two sections relies on three elements: a conceptual claim about the nature of decisions, a theory of the cognitive architecture of deciding (among other things), and a view of what it would take for a propositional attitude to become conscious. Although I represent these elements within a single argument – what I call the Causal Role Argument – I postpone discussion of the conceptual claim until section 5.

We can represent the Causal Role Argument as follows.

1 A decision to act “should issue in motor instructions without the intervention of any further practical reasoning” (2009a, p. 133).

2 No conscious events⁷ play this causal role.

C No practical decision is a conscious event.

In this section I focus on premise 2. Carruthers defends this claim by appeal to a theory of the cognitive architecture of deciding, as well as a view of what it would take for a propositional attitude to become conscious. According to the former, practical decisions are handled largely by a decision-making system which takes as inputs the outputs of a number of systems, for example perceptual systems, belief- and desire-generating systems, and the mindreading system (2010, p. 79). Importantly, the operations of the decision-making system are screened off from consciousness. To understand why, we have to understand what it would take for a propositional

⁷ By ‘conscious event’ I mean an event of which an agent is aware. This usage mimics that concerning awareness of decisions, discussed in section 2.

attitude to become conscious. On Carruthers' view, a propositional attitude PA would become conscious if the system responsible for forming PA was directly connected to a system that could introspect PA – in this case, the mindreading system. The intuition here is that introspection is necessary for consciousness: it is “a presupposition of our common-sense conception of consciousness that our access to our conscious mental states is *not* interpretative . . .” (2007, p. 201). In the case of practical deciding, Carruthers argues that no causal pathway exists from the outputs of the decision-making system to the mindreading system, thus the mindreading system can only self-attribute decisions by way of interpretive means – e.g., by way of particular beliefs about the agent's physical or cognitive environment, or current circumstances.

An example starring Agent Z will facilitate discussion. Z desires to leave work early, believes she can if she decides to, and is uncertain about whether she should. Z comes to believe that leaving early would be best. After forming this belief, Z remains undecided for a few more seconds as she reviews again reasons in favor of staying at work. After doing so she says to herself, ‘I'm leaving,’ and she stands to go.

Here is a plausible interpretation of this story. Concurrent with Z uttering ‘I'm leaving’ in inner speech, Z consciously performs a mental action of assent to an action-plan which involves her going home early. On this interpretation, the action of assent forms an intention to go home, which issues in motor instructions without further practical deliberation (cf. Adams and Mele, 1992).

In virtue of his claim that no causal pathway exists from the outputs of the decision-making system to the mindreading system, Carruthers denies that such an interpretation could apply to humans. Is such a denial plausible for cases of practical deciding? A large part of Carruthers' case depends on interpretations of the confabulation literature critiqued above. This

is bad news for the denial. Worse, an alternative cognitive architecture for practical deciding gives us some reason to accept the above interpretation. This is the architecture outlined by so-called dual-process theories of cognition. Such theories distinguish between system 1 and system 2 reasoning, where system 1 processes are typically fast, automatic, unconscious, and effortless, and system 2 processes are typically slow, controlled, conscious, and effortful (see Evans & Frankish, 2009). It has been suggested by such theorists that various conscious activities, including conscious decisions, are likely to be carried out by system 2 processes. Frankish (2009), for example, suggests that inner speech utterances might qualify as genuine decisions, insofar as such utterances “terminate reasoning processes *at the higher level and on the relevant topic*” (2009, p. 146).

Carruthers is familiar with dual-process theories, and is willing in general to accept the story they tell about cognition. But he argues that attention to the details supports, rather than undermines, his cognitive architecture. Consider again *Z*'s process of deciding to go home early. How *exactly* is her intention formed? Drawing on Frankish's (2004) dual-process account, Carruthers suggests that *Z*'s conscious assent will not lead to action until further system 1 (unconscious) events of practical reasoning have taken place, for example *believing that I have decided to do Q*, and *wanting to do what I have decided*. Thus, “while the act of saying to myself, ‘I shall do Q’ is conscious, and does play a causal role in the production of the behavior of doing Q, it doesn't have the causal role characteristic of a genuine decision to do Q. And so it turns out, after all, that there is no such thing as conscious deciding” (2007, p. 209).

There are reasons to doubt, however, that this is a plausible construal of a dual-process account of practical deciding. Consider what must take place in between Agent *Z*'s uttering ‘I'm leaving’ and her standing to go. On Carruthers' construal, by way of some presumably

collaborative process between her mindreading and belief-generating systems, Z must come to believe that she has decided to leave, at which point this belief will be sent to the practical reasoning system, which presumably houses a standing desire to do what has been decided. But in the case of practical deciding, both propositional attitudes are suspect.

Consider the desire ‘to do what I have intended.’ Leading accounts of the nature of intentions see them as both distinct from beliefs and desires, and more intimately tied to action.⁸ Intentions contain an executive component that other attitudes do not. For example, one plausible account of the nature of intentions views them as executive attitudes towards action-plans, where these plans constitute the attitude’s representational content (Mele, 1992, pp. 197-227). Notice that on such an account, intentions contain attitudes towards plans which are sufficient to initiate action. Once an agent forms an intention by deciding to act, there is no need for a further desire ‘to do what I have intended.’ The existence of the desire in question is thus highly suspect. The same is true of the belief ‘that I have decided to do Q.’ In confabulatory cases, of course, it is plausible that such beliefs will be formed. However, such beliefs will usually be formed after the fact. If beliefs about what I have decided are a crucial part of an account of actual intention formation, they seem not confabulatory, but merely odd. For it is the function of the practical reasoning system to make decisions. It is thus difficult to see why the decision-making system would have much use for prior attributions of decisions emanating from the mindreading or belief-generating system.

A further problem for Carruthers’ construal is that it is not clearly in line with Frankish’s (2004) dual-process account of cognition. Frankish’s effort is directed primarily to developing an account of the role of system 1 and system 2 processes for phenomena like judgments and the

⁸ It is true that some philosophers have attempted to analyze intentions in terms of propositional attitudes like belief and desire (e.g., Davis 1984). But such attempts have not succeeded.

acceptance of premises for use in reasoning. But an account of judgment is different from an account of practical decision. This becomes clear when Frankish briefly develops a dual-process account of practical decision. First, Frankish explicitly restricts his attention to the formation of distal intentions – intentions to do Q *later* (2004, p. 209). Proximal intentions – e.g., Agent Z’s intention to go home *now* – are not discussed. Second, Frankish posits the need for the aforementioned belief ‘that I have decided to do Q’ and desire ‘to do what I have intended’ not to explain the generation of action – which is the primary function of proximal intentions – but rather to explain the role distal intentions play in further cognitive processes (e.g., planning) as well as phenomena like weakness of will and self-deception. On Frankish’s account, an agent can avoid weakness of will if her decision to Q later is accompanied by a commitment (or policy) to maintain Q as a goal as well as a desire to do Q that is greater than any competing desires that arise before Q is done. (Note that the desire to do Q need not be formed by any conscious utterance – the desire to do Q is plausibly a part of the utterance’s causal history, and plausibly persists until Q is done.) This is an interesting proposal, but we do not have space to worry about it here. The relevant point is that on Frankish’s account the existence of commitments and desires is only necessary to explain the *planning* function of *distal* intentions. Since proximal intentions have no essential planning function, it is implausible to posit such elements in an account of proximal intention formation.

Carruthers’ cognitive architecture for practical deciding does not easily fit within a dual-process account of practical deciding. As a result, it generates implausible explanations of the relation between conscious episodes of reasoning, conscious utterances, and the decisions to which such utterances seem tied. I conclude that premise 2 of the Causal Role Argument should be rejected.

5. Real Conscious Decisions

Without premise 2, the Causal Role Argument is unsuccessful. Even so, that argument's first premise states a seeming conceptual truth about the nature of practical decisions. It will be instructive to consider this premise separately. Carruthers claims that a decision to act "should issue in motor instructions without the intervention of any further practical reasoning" (2009a, p. 133). Although here Carruthers frames the issue in terms of motor instructions, this might be a red herring. Practical decisions sometimes concern mental operations not constituted by motor instructions. Consider Jeff conversing with Jane at a party. Though Jeff is invested in Jane's story, he is distracted when someone across the room utters his name. Now Jeff is uncertain about what to do – should he listen to Jane or attend to the conversation across the room? Jeff figures it is best to be polite and so he decides to ignore the conversation across the room, and to attend to Jane. Jeff's decision need not issue in motor instructions directly. However, Jeff's decision arguably must issue in relevant instructions directly – e.g., instructions to the attentional system.⁹ In what follows, then, I will not put weight on the issuance of motor instructions.

What is crucial for Carruthers is the following: "a genuine decision must be the last *deliberative* mental event in the causal chain that leads to the action" (2007, p. 209). As we saw in section 4, on Carruthers' model conscious actions of assent are not the last deliberative events. They thus fail to qualify as conscious decisions.

⁹ Related to this issue, one finds this puzzling admission by Carruthers: ". . . *sometimes* a decision to direct attention at a stimulus or type of stimulus can be conscious (as when I remind myself when doing a jigsaw puzzle, 'I must pay attention to the shape'). But the intervening processes that begin from a thought of this sort and issue in a globally broadcast perception or image won't themselves be conscious ones" (2009b, pp. 114-115). The point about later events being conscious or not seems tangential. If we can make conscious decisions about where to direct attention why should we accept that we never make conscious decisions about actions of other kinds?

I have given reasons to reject Carruthers' model of action initiation. Suppose, however, that future research vindicates the model. We discover that conscious events cannot produce intentions or initiate motor instructions without the assistance of further non-conscious practical reasoning processes. When you consciously assent to an action plan, A, no intention is yet formed. Rather, non-conscious processes take this event of conscious assent as input, and *interpret* it. In most cases, these non-conscious processes conclude that you have committed to A-ing – a conclusion that results in the formation of the belief that you will A. Then these processes integrate this belief with your standing non-conscious desire to execute your commitments. At this moment, an intention is formed and, if your action-plan was to A immediately, motor instructions are sent.

Could we consider such a process a conscious decision? The answer depends partially on the relationship between conscious assent and subsequent action. If we possessed rogue non-conscious processes, which commonly contradicted conscious assent by forming contradictory beliefs and thus actions, then the answer would be no. But we can consider a weaker possibility, by asking: do non-conscious processes ever undermine conscious assent, thereby generating contrary action? I think that they do. Some such cases are discussed in the literature on weakness of will. What we want to know is whether the undermining of conscious assent in *some* cases renders plausible the claim that we *never* consciously decide to act. I doubt that it does, for the following reason. In some contexts, conscious assent reliably leads to action. How does it do so? On the option we are considering, the route will not be one of direct causation. Perhaps conscious assent to A-ing simply raises the probability that one will A. If so, then our choice to treat conscious assent as a conscious decision will depend partially on how reliably conscious assent produces corresponding action.

Consider the following case. When you assent to A-ing, your action plan is this: ‘get off the couch in five seconds.’ You assent to this plan by considering the prospect of getting off the couch, and by saying to yourself, with this plan in mind, ‘sounds good.’ You then begin the countdown. Let’s say that this conscious assent raises the probability that your non-conscious processes will form an intention to get off the couch to .9 – as it happens, this is the conclusion reached. After you finish the countdown, motor instructions are sent and you get off the couch. We are granting that the conscious event does not terminate practical reasoning. But it seems we want the answer to a different question: are the further non-conscious practical reasoning events *in charge*? It is plausible that it is rather the conscious event that plays the executive role – after all, it raised the probability that your non-conscious processes would come to the interpretation they did to .9. Is it plausible that the *whole* event deserves to be called a *conscious* decision?

Carruthers is set against this possibility. He argues that, given his architecture for deciding, “There are introspectable events that sometimes give rise to judgments and decisions (items of inner speech or other forms of imagery); but these aren’t, themselves, the judgments and decisions. And there are, of course, such things as judgments and decisions; but these aren’t introspectable” (2010, p. 102). The thought here is that the *real* decision takes place the moment the intention is formed. And indeed this is what Mele (2009) holds as well. Given this concept of deciding, the process described above would not qualify as a conscious decision.

Notice, however, that conscious assent can retain a crucial role, even if Carruthers’ architecture gives the correct causal story of practical deciding. Indeed, if we assume that the process described above is often actual, decisions come to resemble mental episodes like forgetting and remembering. Agents can try to bring it about that they forget or remember something, and sometimes they are successful. Similarly, if Carruthers’ architecture is correct it

seems that agents can try to bring it about that they decide to A by consciously assenting to A-ing. Unlike forgetting and remembering, however, such tryings would seem to bear a particularly intimate connection to intention formation. Tell yourself (and mean it) to flex your left hand now, your right hand now, or to do neither. If you are like me, you did what you told yourself to do. Further, you could repeat this experiment in different circumstances with similar results. It is plausible to think that you could because your conscious assent to A-ing bears an intimate connection to your formation of an intention to A.

If we discovered that Carruthers' architecture was correct, and we discovered that an agent's conscious assent to A-ing significantly raised the probability of an agent's forming an intention to A, this might warrant *revision* of the concept of deciding. Here is a suggested revision. Consciously deciding does not require an agent to be conscious of the action of intention formation the whole time the action was being performed. Rather, the agent need be conscious only of the assent which reliably initiates the process of deciding. It seems that in certain situations, conscious assent to A-ing does reliably initiate A-ing, presumably by (at least) initiating the process of deciding to A. In some of these situations, as we have seen, important details of the action-plan get sorted through and consciously assented to before post-conscious processes are able to do their work. Given this, even if Carruthers' architecture is vindicated by future research we might plausibly maintain that humans sometimes consciously decide to act.

6. Conclusion

In this paper, I have considered Carruthers' case in support of the claim that we never consciously decide to act. This case appeals to empirical and conceptual considerations. I have argued that, on close consideration, the empirical data fail to support the claim. And I have

argued that *given* Carruthers' architecture for deciding, the conceptual considerations lack force. I conclude that Carruthers' claim that we never consciously decide to act is established by none of his arguments. I stress that nothing above establishes the claim that we do consciously decide to act. Indeed, one salutary effect of Carruthers' arguments – and of my counter-arguments, I hope – is to focus attention on the importance of certain empirical and conceptual considerations to the issue of conscious deciding. It seems we have some work to do before we can conclude with much confidence anything about conscious deciding.

References

- Adams, F. & Mele, A. (1992). The intention/volition debate. *Canadian Journal of Philosophy*, 22, 323-338.
- Bach, K. (1980). Actions are not events. *Mind*, 89, 114-120.
- Brasil-Neto, J., Pascual-Leone, A., Valls-Sole', J., Cohen, L. & Hallett, M. (1992). Focal transcranial magnetic stimulation and response bias in a forced choice task. *Journal of Neurology, Neurosurgery, and Psychiatry*, 55, 964-66.
- Carruthers, P. (2006). *The Architecture of the Mind*, Clarendon Press, Oxford University Press.
- _____ (2007). The illusion of conscious will. *Synthese*, 159, 197-213.
- _____ (2009a). Action-awareness and the active mind. *Philosophical Papers*, 38, 133-156.
- _____ (2009b). An architecture for dual reasoning. In J. St. B. T. Evans & K. Frankish (Eds.). *In Two Minds: Dual Processes and Beyond* (pp. 109-128). Oxford University Press.

- _____ (2009c). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32, 121-138.
- _____ (2009d). Mindreading underlies metacognition. *Behavioral and Brain Sciences*, 32, 164-176.
- _____ (2009e). Simulation and the first person. *Philosophical Studies*, 144, 467-475.
- _____ (2010). Introspection: divided and partly eliminated. *Philosophy and Phenomenological Research*, 80, 76-111.
- Clarke, R. (2003). *Libertarian Accounts of Free Will*. Oxford University Press.
- Davis, W. A. (1984). A causal theory of intending. *American Philosophical Quarterly*, 21, 43-54.
- Evans, J. St. B.T. & Frankish, K. (Eds.). (2009). *In Two Minds: Dual Processes and Beyond*. Oxford University Press.
- Fiala, B. & Nichols, S. (2009). Confabulation, confidence, and introspection. *Behavioral and Brain Sciences*, 32, 144-145.
- Frankfurt, H. (1988). *The Importance of What We Care About*. Cambridge University Press.
- Frankish, K. (2004). *Mind and Supermind*. Cambridge University Press.
- Frankish, K. (2009). How we know our conscious minds: Introspective access to conscious thoughts. *Behavioral and Brain Sciences*, 32, 145-146.
- Geach, P. (1957). *Mental Acts: Their Content and Their Objects*. New York, Humanities Press.
- Kane, R. (1996). *The Significance of Free Will*. Oxford University Press.
- McCann, H. (1986). Intrinsic Intentionality. *Theory and Decision*, 20, 247-273.
- Mele, A. (1992). *The Springs of Action*. Oxford University Press.
- _____ (2003). *Motivation and Agency*, Oxford University Press.

- _____ (2005). Decisions, intentions, and free will. *Midwest Studies in Philosophy*, 24, 146-162.
- _____ (2009). *Effective Intentions: The Power of Conscious Will*. Oxford University Press.
- _____ (2010). Conscious deciding and the science of free will. In R.F. Baumeister, A.R. Mele & K.D. Vohs (Eds.) *Free Will and Consciousness: How Might They Work?*, Oxford University Press, 43-65.
- Milner, A. D. and Goodale, M. A. (2006). *The Visual Brain in Action* (2nd ed.), Oxford University Press.
- Nisbett, R. & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-95.
- O'Connor, T. (2009). Degrees of Freedom. *Philosophical Explorations*, 12 (2), 119-125.
- O'Shaughnessy, B. (1980). *The Will* (vol. 2). Cambridge University Press.
- Schroeder, T. (2006). Propositional Attitudes. *Philosophy Compass*, 1(1), 65-73.
- Searle, J. (2001). *Rationality in Action*. MIT Press.
- Selen, L.P.J., Shadlen, M.N. & Wolpert, D.M. (2012). Deliberation in the motor system: Reflex gains track evolving evidence leading to a decision. *The Journal of Neuroscience*, 32(7), 2276-2286.
- Tye, M. (1995). *Ten Problems of Consciousness*. MIT Press.
- Wegner, D. M. (2003). *The Illusion of Conscious Will*. MIT Press.
- Wegner, D. M. & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-492.
- Wilson, T. D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. The Belknap Press of Harvard University Press.