To appear in S. Gallagher (Ed.), *The Oxford Handbook of the Self*. Oxford: Oxford University Press, forthcoming..

Self-Agency

Elisabeth Pacherie

Institut Jean Nicod, Paris

EHESS, DEC-ENS, CNRS, Paris, France

pacherie@ens.fr

Word count: 10662

1. Introduction

We are perceivers, we are thinkers, and we are also agents, bringing about physical events, such as bodily movements and their consequences. What we do tells us, and others, a lot about who we are. On the one hand, who we are determines what we do. On the other hand, acting is also a process of self-discovery and self-shaping. Pivotal to this mutual shaping of self and agency is the sense of agency, or agentive self-awareness, i.e., the sense that one is the agent of an action.[1]

As folk, we appear to be deeply wedded to a conception of self-agency according to which consciousness plays a pervasive role prior to acting, while acting, and after one has acted. Here's how the story goes. *Conscious* deliberation on the basis of our *conscious* beliefs and desires yields a *conscious* decision to pursue a certain *conscious* goal, leading to the formation of a *conscious* intention or volition to realize that goal. Our *conscious* intention in turn causes our action, by *consciously* initiating and *consciously* controlling it. While acting we

[1] In this paper, I will use 'sense of agency' and 'agentive self-awareness' as synonyms.

experience our *conscious* intention as causing our action and on that experiential basis, we are able to judge immediately after acting that we were the agent of the action. On that story, the causation of action by conscious mental states and the sense of agency for actions are but two sides of the same coin. Although not uncontested, this folk-psychological picture has been endorsed by many philosophers. Typically, their qualms have concerned less the role attributed to consciousness than the rather prodigal ontology of mental states the folk is content to entertain (Davidson, 1980) and the difficulties involved in making room for mental causation in a physicalist framework – the infamous exclusion problem (Kim, 1998). The apparently obvious link between the causation of action by conscious mental states and the sense of agency may also explain why until recently the sense of agency was a topic rather neglected by philosophers of action and philosophers of consciousness alike.

Things have changed, however, and of late the sense of agency has regained its place in the agenda of philosophers and scientists alike.[2] This recent explosion of interest in the topic is due in a large part to the development of psychological and neuroscientific methods that have made the phenomenology of action an object of empirical investigation and yielded results that challenge the received wisdom. On the one hand, empirical research done over the last two decades provides evidence that the role of consciousness in action production is far less pervasive than our folk-conception would have it. Many of our actions are selected, carried out and controlled non-consciously, with conscious states playing at best a role in non-default modes of action control (such as trouble-shooting) and being at worst mere epiphenomenal by-products of something else that does the causal work. On the other hand, cognitive scientists have also developed models of how the sense of agency is generated, exploring a number of potential cues to agency and proposing several different mechanisms, ranging from high-level cognitive mechanisms to low-level sensorimotor cues.

[2] Useful collections of papers can be found in Roessler and Eilan (eds.)(2003), Sebanz and Prinz (eds.) (2006), Pockett et al (eds.) (2006), and Siegel (ed.) (2007).

This recent empirical work has done a lot to rekindle the interest of philosophers in these issues by suggesting that they are far more complex than previously thought. First, the sense of agency itself appears to be a complex, multifaceted phenomenon, raising questions about how its various aspects are related, what their content and structure is, to what extent they are dissociable, and whether some are more basic than others. Second, the relationship between sense of agency and conscious mental causation of action appears to be much less straightforward than the two-sides-of-the-same-coin story would have it. To the extent that some of the mechanisms involved in the generation of the sense of agency are only indirectly related to the mechanisms involved in action production, sense of agency and agency may dissociate. We sometimes have a sense of agency for actions we did not actually perform or did not consciously intend, and conversely we may lack a sense of agency for actions we did consciously intend and actually performed. This has led some to claim that the conscious will is an illusion. It isn't entirely clear what exactly this claim is supposed to mean and whether it is warranted. Certainly, however, the existence of such dissociations invites us to reconsider our conception of self-agency, of the link between self-agency and conscious mental causation, and of the role the sense of agency plays within agency.

In order to introduce readers to the new terms of the debate over self-agency, I start with a brief survey of recent models of how and where in the cognitive architecture the sense of agency is generated. I also argue that these models should be seen as complementary rather than as rivals. Building on that basis, I then turn to issues concerning the contents and structure of states of agentive self-awareness. Finally, I discuss the relation between agency and sense of agency. Is the sense of agency a reliable indicator of actual agency? Does it represent agency as it really is? What role does the sense of agency play in agency?

2. How is the sense of agency generated?

Empirical research on agency has explored a number of potential cues to agency, and different cognitive models for agency have been proposed, ranging from high-level cognitive mechanisms to low-level sensorimotor mechanisms.

Cognitive cues and high-level mechanisms

Some authors focus on high-level cognitive mechanisms and invoke a "central" interpretive system to explain our awareness of our own agency. According to this approach, the sense of agency is subserved by a holistic mechanism that is concerned with narrative self-understanding. Our sense of what, if anything, we are up to, is based on the operations of a high-level integrative process that draws on the agent's self-conception and tries to put the best spin on things that it can. Such a conception has strong Dennettian overtones. We turn Dennett's intentional stance inwards, and treat ourselves as entities whose behavior needs to be made sense of in light of an implicit theory of ideal agency.

Many authors have expressed some sympathy with, and in some cases whole-hearted commitment to, the narrative approach. Interpreting split-brain studies in light of Dennettian (1992) themes concerning the role of narrative in self-interpretation, Roser and Gazzaniga (2004; 2006) have argued that the left hemisphere contains an interpreter, whose job it is to make sense of the agent's own behavior. The psychiatrist Louis Sass has suggested that schizophrenic patients with delusions of alien control no longer feel as though they are in control of their actions because "particular thoughts and actions may not make sense in relation to the whole" (1992: 214). Developing Sass's proposal, Stephens and Graham suggest that a "subject's sense of agency regarding episodes in her psychological history might depend on her ability to integrate them into her larger picture of herself" (2000: 161). Peter Carruthers suggests that "our awareness of our own will results from turning our mind-

reading capacities upon themselves, and coming up with the best interpretation of the information that is available to it—where this information doesn't include those acts of deciding themselves, but only the causes and effects of those events." (2007: 199).

Holistic themes also play an important role in Daniel Wegner's influential treatment of agentive self-awareness. On the one hand, Wegner argues that the sense of agency is typically inferred from the existence of a match between a prior thought and an observed action:

> The experience of consciously willing our actions seems to arise primarily when we believe our thoughts have caused our actions. This happens when we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. […] In essence, the theory suggests that we experience ourselves as agents who cause our actions when our minds provide us with previews of the actions that turn out to be accurate when we observe the actions that ensue. (Wegner, 2005: 23)

On the other hand, he also notes that we perform many actions without the benefit of such previews and suggests that "Even when we didn't know what we were doing in advance, we may trust our theory that we consciously will our actions and so find ourselves forced to imagine or confabulate memories of "prior" consistent thoughts." (Wegner, 2002: 146)

A wide array of evidence can be marshaled in support of this high-level account. When young children happen to achieve a goal by luck, they will say that they had intended the action that yielded that goal all along (Phillips, Baron-Cohen, & Rutter, 1998). Proponents of this approach have also appealed to studies involving patients with brain damage. Patients with anosognosia for hemiplegia say that they are currently raising their arm when, in fact, their arm has not moved. When it is pointed out to the patient that his arm has not moved, he may confabulate an excuse for his inertia (Feinberg, Roane, & Ali, 2000) . Split-brain subjects are

prone to confabulate accounts of actions that are generated by their right hemisphere (Gazzaniga & LeDoux, 1978). Data from subjects in altered states of consciousness also support the narrative approach. For example, bizarre behaviors performed in response to hypnotic suggestion are often accompanied by elaborate rationalizations and confabulation on the part of the agent (Moll, 1889). Finally, this approach derives support from a number of laboratory studies with normal subjects, in which it has been shown that the sense of agency can be modulated by priming and various contextual parameters (Aarts, Custers, & Wegner, 2005; Wegner, Sparrow, & Winerman, 2004; Wegner & Wheatley, 1999).

Sensorimotor cues and low-level mechanisms

In contrast to this very high-level approach, a number of researchers have proposed that the monitoring of action execution is crucial for agency and that the sense of agency is generated by low-level mechanisms that exploit performance-related sensorimotor cues.

One possibility is that efferent signals sent to the motor system while implementing an intention provide such cues. Tsakiris and colleagues have proposed that efferent signals are used to generate accurate temporal and kinematic predictions about how and when particular body parts should move (Tsakiris & Haggard, 2005; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005; Tsakiris, Prabhu, & Haggard, 2006). In support of that claim, they demonstrated that self-recognition of one's own bodily movements crucially depends on efferent signals.

Another line of evidence for the role of efferent signals in generating a sense of agency involves ‘intentional binding’, a phenomenon in which self-produced movements and their effects are perceived as being closer together in subjective time than they actually are (Haggard & Clark, 2003; Haggard, Clark, & Kalogeras, 2002). More specifically, when a voluntary act (e.g., a button press) causes an effect (e.g., a tone), the action is perceived by the agent as having occurred later than it did, and the effect is perceived as having occurred

earlier. In contrast, when similar movements and auditory effects occur involuntarily rather than voluntarily, the binding effect is reversed and cause and effect are perceived as further apart in time than they actually are. Haggard suggests that intentional binding is best explained in terms of predictive mechanisms of action control: it depends on efferent signals since it does not occur with passive movements and it causes anticipatory awareness of action effects, a shift that suggests prediction. On this predictive account, the sense of agency would be constructed at the time of the action itself, as an immediate by-product of the motor control circuits that generate and control the physical movement itself.

Another mechanism appeals to internal forward models used for action control (Blakemore & Frith, 2003; Frith, Blakemore, & Wolpert, 2000a, 2000b). According to this proposal, forward models are fed an efference copy of actual motor commands and compute estimates of the sensory consequences of the ensuing movements. The predicted sensory consequences are compared with actual sensory feedback (reafferences) When there is a match between predicted and actual state, the comparator sends a signal to the effect that the sensory changes are self-generated, and when there is no match (or an insufficiently robust match), sensory changes are coded as externally caused. Indirect evidence for this model comes from studies demonstrating that discrepancies between predictions and sensory reafferences affect tactile sensations (Blakemore, Wolpert, & Frith, 1998; Blakemore, Wolpert, & Frith, 2000) and visual perception of one's own actions (Leube et al., 2003) et al., 2003). Direct evidence is also provided by studies demonstrating that agency is gradually reduced as these discrepancies increase due to spatial deviations and temporal delays (Fourneret & Jeannerod, 1998; Knoblich & Kircher, 2004; Knoblich, Stottmeister, & Kircher, 2004; Leube et al., 2003; Sato & Yasuda, 2005; van den Bos & Jeannerod, 2002);

Perceptual cues and intermediate level mechanisms

However, as several authors have pointed out (Gallagher, 2007; Knoblich & Repp, 2009; Pacherie, 2008), the results of some of these studies are open to alternative interpretations in terms of perceptual rather than sensorimotor cues. It is well known that we have little awareness of the proprioceptive feedback associated with movements or even of the corrections we make during goal directed movements (De Vignemont, Tsakiris, & Haggard, 2006; Fourneret & Jeannerod, 1998). Indeed, passive movements are associated with more activity in the secondary somatosensory cortex than active movements (Weiller et al., 1996) Frith (2005) even suggests that lack of proprioceptive experience may be one indicator that one is performing a voluntary act. The vast majority of our actions aim at producing effects in the environment and we normally attend to the perceptual effects of our movements rather than to the movements themselves. It may therefore be that perceptual cues rather than sensorimotor cues are crucial to the sense of agency. Direct evidence for this view comes from an experiment of Fourneret and Jeannerod (1998) where subjects are instructed to move a stylus on a graphic tablet on a straight line to a visual target. Subjects cannot see their drawing hand, only its trajectory being visible as a line on a computer screen. However, the experimenter introduces a directional bias electronically so that the visible trajectory no longer corresponds to that of the hand. When the bias is small (< 14°) subjects make automatic adjustments of their hand movements to reach the target but remain unaware that they are making these corrections. It is with larger biases that subjects become aware of a discrepancy and begin to use conscious monitoring of their hand movement to correct for it and to reach the target. These results suggest that although discrepancies between predicted and actual sensory feedback are detected at some level since they are used to make appropriate corrections of the hand movement, they do not influence the sense of agency. Rather, their sense of agency for the action seems to rely mostly on a comparison of the predicted and actual perceptual consequences of their action. As long as the trajectory seen on

the screen matches sufficiently well the predicted trajectory, proprioceptive information is ignored.

Further evidence that perceptual cues may contribute more to the sense of agency than sensorimotor cues comes from pathologies (Jeannerod, 2009). For instance, schizophrenic patients are impaired in explicitly judging whether they are in control of perceptual events but not impaired in automatically compensating for sensorimotor transformations between their movements and the resulting perceptual events (Fourneret et al., 2002). Frontal patients, like schizophrenic patients, have a preserved automatic sensorimotor control, contrasting with impaired action awareness and conscious monitoring (Slachevsky et al., 2003).

Toward an integrated model of agentive self-awareness

All the models I briefly reviewed share a core idea. They appeal to a principle of congruence between anticipated outcome and actual outcome. Where they differ is on whether the cues used are primarily cognitive, perceptual or sensorimotor and on how closely they are related to action production mechanisms and intentional processes. There is now, however, a growing consensus that these different models should be seen as complementary rather than as rivals and that the sense of agency relies on a multiplicity of cues coming from different sources (Bayne & Pacherie, 2007; Gallagher, 2007; Knoblich & Repp, 2009; Pacherie, 2008; Sato, 2009; Synofzik, Vosgerau, & Newen, 2008).

Thus, the conceptual framework I proposed (Pacherie, 2008) distinguishes between three hierarchically ordered intentional levels: (1) distal intentions, where the action to be performed (i.e. goals and means) is specified in cognitive terms, (2) proximal intentions, where it is specified in actional-perceptual terms, that is terms of the action schemas to be implemented and the perceptual events that will occur as a consequence, and (3) motor

intentions where it is specified in sensorimotor terms. Comparisons of desired, predicted and actual states at each of these three levels provide different cues to agency.

This framework, as well as other similar integrative frameworks, leaves open a number of questions. What is the relative weight of cues on each of the three levels with regard to experiencing agency? To what extent can cognitive expectations overrule perceptual and sensorimotor evidence? To what extent can the relative weight of different agency cues be modulated by the nature of the task, by the attentional state of the agent, or by the agent's level of expertise? To answer those questions empirical investigations are needed. By themselves, however, such integrative frameworks may also help allay some of the controversies and perplexities surrounding the content, mode, structure and reliability of agentive self-awareness.

3. Content, mode, and structure of agentive self-awareness

To have a sense of agency for an action is to be aware of oneself as the agent of the action. But what form(s) does this awareness take? To answer this question, several dimensions of agentive self-awareness must be considered. A first dimension along which states of agentive self-awareness can be located concerns their representational contents. A second, closely related, dimension concerns their mode. A third concern their structure, what the direction of fit is supposed to be between theses states and the world. A fourth important dimension is the temporal dimension: states of agentive self-awareness may differ according to whether one is about to act, one is in the process of acting, or one has already completed the action. As we will see, these dimensions are not strictly orthogonal, rather they interact in a variety of ways. I start with agentive content and mode.

Agentive content and mode

Agentive contents can range from the ‘thin’ to the ‘thick’. At the thin end of the spectrum, one can experience oneself as acting and as acting more or less effortfully. For instance, one can be aware of one's bodily movements as active rather than passive. Moving up the spectrum, agentive content can include not merely the representation of a movement as one’s own action, but also a representation of its effects in the world as effects one brought about. Moving even further up, agentive content can include a representation of the kind of action one is performing and one’s reasons for performing it. One can be aware of oneself as opening a door, and as opening a door in order to (say) leave a building (as opposed to showing someone how to open the door). At the thicker end of the spectrum, some authors argue that agentive contents represent our mental states as causing our actions (Hohwy, 2004). They might say, for instance, that I am aware not just that I am opening the door in order to leave the building, but that I am aware my opening the door as caused by my intention to leave the building. Or, in the self-referential variant endorsed by Searle (1983) and Mossel (2005), that states of agentive self-awareness represent themselves as causing the target action.

Not all theoricians agree that agentive contents can display such richness. Some hold onto an austere conception that only allows for thin contents. One way to make sense of their reticence is in terms of a second important dimension of agentive self-awareness. Bayne & Pacherie (2007) distinguish between agentive judgments and agentive experiences. Gallagher (2007) draws a similar distinction between two levels of the sense of agency, contrasting them in terms of first-order phenomenal experience of agency and higher-order, reflective attribution. Some states of agentive self-awareness take a judgmental (or doxastic) form. Normally, one’s awareness of the kinds of actions that one is performing —making a cup of coffee, writing a paper—takes the form of belief. Yet, the ‘vehicles’ of agentive self-awareness are often more primitive than judgments. Consider what it is like to push a door

open. One might judge that one is the agent of this action, but this judgment is not the only way in which one's own agency is manifested to oneself. Instead, one experiences oneself as the agent of this action. Such states are no more judgments than are visual experiences of the scene in front of one or proprioceptive experiences of the current position of one's limbs..

It seems that proponents of an austere conception of agentive content have agentive experiences in mind. It is very much an open question how rich the contents of agentive experience can be. Arguably, there are restrictions on the kind of properties than can be experientially encoded. A strong case can be made for thinking that the contents of agentive experience can go beyond merely representing oneself as actively moving, but also include information about the degree of control one has over the movement and the degree to which the action is effortful. But it is less clear to what extent the intentional content of an action or one's aims in performing the action can also be encoded in agentive experience. Insofar as matches and mismatches between perceptual predictions and feedback appear to be important cues to agency, I am inclined to consider that agentive experience can partake of the richness of perceptual experience. Similarly, Gallagher (2007) suggests that the sense of agency as a first-order experience includes not just sensorimotor content linked to bodily movement but also perceptual information linked to the effects of the action in the world.

Whatever the constraints on the contents of agentive experiences, they are no obvious restrictions on cognitive content; almost any property can be represented in cognition. Insofar, as agentive judgments have cognitive contents, they are therefore not obvious limits to their richness. Thus, as long as one is willing to countenance both agentive experiences and agentive judgments as complementary forms of agentive self-awareness, the debate about agentive content shouldn't really be about how rich it is possible for it to be. Rather, the debate is about how this content is distributed between agentive experience and agentive judgment and about how experience and judgment relate.

With respect to the distribution of agentive content, one important issue beyond those already considered concerns the cognitive penetrability of agentive experience. One aspect of agentive experience—our moment-by-moment sense of ourselves as the agents of various movements—is largely based on sensorimotor cues from low-level, comparator-based systems. These low-level systems are largely modular and as such presumably rather impervious to top-down influences. Thus, Fried and colleagues (Fried et al., 1991) reported that electrical stimulation of the supplementary motor area could elicit in their patients a subjective "urge" to perform a movement in the absence of overt motor response. Furthermore, at some sites where these subjective experiences were elicited, stimulation at higher current evoked an overt motor response. Subjects in these experiments had all the reasons to judge that they were not the agents of their movements as they knew these were caused by the electrical stimulation applied to their brain; yet they felt the urge to move and when they actually moved had an experience of actively moving.

Yet, as we saw in section 2, there are reasons to think that perceptual cues may contribute more to agentive experience than sensorimotor cues. Now there's a better case to be made for cognitive penetrability of perceptual content. In one of the classic papers of New Look psychology, Bruner and Goodman (1947) found that poorer children perceived coins as bigger than rich children do. This study suggests that high-level information can affect the contents of visual perception. Similarly, it might be possible for cognitive expectations to exert a top-down influence on agentive experience at least around the margins.

Conversely, one may ask how strong the dependence of agentive judgments on agentive experience is. Agentive judgments are typically grounded in and justified by agentive experiences. In the normal case, we judge that we are the agent of a particular action on the grounds that we enjoy an agentive experience with respect to it; here, our agentive judgments are simply endorsements of our agentive experiences. Correlatively, we normally judge that

we are not the agent of a particular event because we lack an agentive experience with respect to it. In light of this, pathologies of agentive self-awareness in which the agent denies that one of their actions is their own are likely to be grounded in pathologies of agentive experience. The aetiological account of such pathologies would, on this approach, be primarily a matter of accounting for disturbances in the agent's experience of their own agency (Pacherie, Green, & Bayne, 2006).

Yet, our agentive judgments are often more than just endorsements of our agentive experiences. Although my agentive experience might tell me that I performed a certain action, it may not tell me what kind of action it was or what my reasons were for performing it. The job of agentive judgments would then be to provide an interpretation of the action in the light of one's self-conception. They may do so, as Wegner suggests, by either linking the action to some consistent prior thought we had or, if need be, by confabulating reasons for the action.

Furthermore, our agentive judgments are not beholden to our agentive experiences. Not only can we deny that we are the authors of events towards which we have an agentive experience, we can also assert that we are the authors of events for which we lack such an experience. The agent's narrative self-conception cognitive expectations might place rich and substantive constraints on whether or not the contents of agentive experiences are to be accepted. In general, experiential states do not compel assent, and there is no reason to think that matters are any different with respect to agentive experiences. It might be the case that agents evaluate their agentive experience (or lack thereof) in light of their narrative self-conception. Agentive experience for an action may be overridden by certain holistic constraints, in instances where the action makes no sense with regard to our narrative self-conception. Or conversely, we may self-attribute an action despite lacking an experience of agency for that action, if the action makes perfect sense in light of our narrative self-conception. The latter may actually be quite common, as there is reason to think that in many cases the agent won't

have any information about agentive experience to draw on in forming agentive judgments. Such experiences are likely to be labile and short-lived; leaving no trace in long-term memory unless attentional resources are used to probe and consolidate them. Given their short life-span, our agentive experiences may well have been obliterated by the time we make an agentive judgment. Such 'gaps' between agentive experience and agentive judgment might be large enough for processes of narrative reconstruction to exploit, allowing an agent's narrative self-conception to restructure their agentive self-awareness.

Structure and time course

Bayne (2008) distinguishes four possible conceptions of the structure of states of agentive self-awareness. As a first possibility one may think of them as having a descriptive structure. On this view, they are supposed to say how things are, have a mind-to-world direction of fit, a world-to-mind direction of causation and veridicality conditions. Alternatively, one may hold that states of agentive self-awareness have a directive structure, representing how the world is to be changed and having a world-to-mind direction of fit and a mind-to-world direction of causation. On the directive account then, these states would have conditions of satisfaction but not veridicality conditions. The third account holds that states of agentive self-awareness are akin to Millikan's *pushmi-pullyu* representations (Millikan, 1995), having at once a mind-to-world and a world-to-mind direction of fit and thus being both descriptive and directive. According to the fourth and final account, states of agentive self-awareness would lack intentional content altogether, involving only raw phenomenal feels. As pointed out by Bayne, although Searle (1983) did much to put the directive account on the table, the descriptive account seems to be the majority view[3], while the "pushmi-pullyu" and the raw feels accounts have not be developed in any detail.

[3] For recent defenses of the descriptive account, see Bayne (2009) and Proust (2003).

In my view, all these accounts have some merits. Yet, given the complex nature and the many facets of agentive self-awareness, it is unlikely that a single account should hold across the board. Let us first consider the distinction between agentive experience and agentive judgment. Clearly, the only plausible account of the structure of agentive judgments is the descriptive account. Agentive judgments have a mind-to-world direction of fit and veridicality conditions. Thus, it is only when one focuses on agentive experiences that there is room for debate. Second, we should also note that specific temporal constraints apply to states with directive structure. No particular temporal constraints on their contents are imposed on states in virtue of their having a descriptive structure. For instance, my beliefs can be about past, present or future states of affairs. When temporal constraints apply, they do apply not in virtue of the state having a descriptive structure but in virtue of features of its mode (as in regret that can only concern past states of affairs). In contrast, there are temporal constraints on the contents of states with a directive structure. I can intend to do something now or to do something tomorrow, but I cannot intend to do something yesterday. These constraints stem from the fact that these states have a mind-to-world direction of causation and from the time-asymmetry of causation: effects cannot precede their causes. For agentive experiences to have a directive structure, it would therefore be necessary, although not sufficient, that they satisfy those temporal requirements. In other words, it would be necessary that these experiences occur before the state of affairs they are about.

Among the cues contributing to agentive experiences that we discussed in section 2, some but not all satisfy these requirements. Matches between desired and actual (sensory or perceptual) effects as well as between predicted and actual effects are thought to provide cues to self-agency, but for there to be such matches, the actual effects must have been brought about already. So the agentive contents contributed by these cues cannot have a directive structure. There are other cues to self-agency, however, that seem to have the right temporal

characteristics. Efferent signals (Tsakiris & Haggard, 2005; Tsakiris et al., 2005) and matches between desired and predicted effects both provide cues to agency in advance of actual movement, thus tagging the forthcoming event as being "mine". Yet, this does not suffice to vindicate the directive account. There should also be a mind-to world direction of causation. Obviously, efferent signals contribute to the production of the action, but what about matches between desired and predicted states? In computational models of the motor system (Miall & Wolpert, 1996) the comparison of predicted and desired states is seen as final check before the gates to action execution are opened, suggesting indeed that signals arising from that comparison process control the gates to action execution. The signals would thus underlie the sense that one is initiating an action.

So we have signals that appear to both play a role in action production and function as cues to self-agency. Is the directive account thereby vindicated? Not quite. The phrase " X is a cue to self-agency" can be understood in two different ways: as meaning that X is a constituent of agentive experience (the constitutive reading) or as meaning that X plays a causal role in bringing about the agentive experience (the causal reading). Only the constitutive reading is consistent with the directive account, the causal reading would belong with a descriptive account. I find it hard to tell how much this is an empirical as opposed to a conceptual issue. Following Bayne (2008), we could say that the constitutive reading would rule out the possibility of a person experiencing herself as initiating an action she is not actually initiating, whereas the causal reading would allow it, suggesting this is at least in part an empirical issue. In practice, however, it may prove extremely difficult to completely separate out the various components of agentive experience and thus test for such a dissociation.

Finally, even if we opt for the constitutive reading, we are not yet out of the woods. The directive account, like the descriptivist account, is representationalist in spirit. But do the relevant signals carry intentional content accessible to consciousness? Pure efferent signals

may be thought to be responsible for raw feelings of activity and effort, amenable to a raw feels account rather than to the directive account. Signals arising from the comparison of desired and predicted states present us with better prospects, for obviously the states compared carry intentional content. Yet it is unclear whether we have access to the contents of the states compared or simply to the result of their comparison. As we saw in section 2, empirical evidence indicates that we are largely unaware of the nature of our movements. This suggests that at the sensorimotor level at least, we have no conscious access to the contents of sensorimotor representations of desired and predicted states or to contents encoding the nature of the possible discrepancies between them. Rather we may simply have access to information that there is or there isn't a match. The agentive experience would then have very thin content indeed, something we may paraphrase as "Ok, let's go" or "Wait, not ready". Comparisons of desired and predictive effects made at the proximal level may however provide more glamorous agentive content, for here empirical evidence suggest that we normally have precise conscious expectations regarding the perceptual events our action is to bring about. So something like "Let there be light!" may enter the content of our agentive experience when about to press a switch.

Experiences with agentive content based on predictive signals may then constitute the most convincing case for the directive account. However, given the dual role predictions play in motor control, they may also be amenable to a pushmi-pullyu interpretation. On the one hand, comparisons of desired with predicted states control the gates to action execution; on the other hand, comparisons of predicted with actual states signal either completion or the need for corrections. With respect to their former function, they seem to have a directive structure, with respect to the latter a descriptive structure. That a single representation might have two opposite directions of fit may appear paradoxal. This air of paradox may be somewhat alleviated if one considers the time-course of these representations. They would seem to have

a directive structure in the early stages of action, but a descriptive structure in its late stages. When all goes well and action execution proceeds without impediment, the intentional content of predictions remain the same, but their direction of fit gradually changes from directive to descriptive.

In a nutshell, each of the four accounts of the structure of agentive self-awareness distinguished by Bayne may have some plausibility for at least some components of agentive self-awareness. The descriptive account is certainly the one with the widest scope. All the components of agentive self-awareness that are judgmental in form have descriptive structure and so do many experiential components of agentive self-awareness. If one considers, however, that agentive experience is not a point-like experience but an experience extended in time and evolving as the action unfolds, one should be open to the possibility that some components of this experience have a directive or a pushmi-pullyu structure. Finally, the brute efferent component of agentive experience may be responsible for raw phenomenal feels of activity and effort.

4. Agency and sense of agency

How reliable is the sense of agency?

This question takes different forms depending on what we take the structure of states of agentive self-awareness to be. If one holds that states of agentive self-awareness have a descriptive structure, the question whether they are reliable becomes the question whether or not they tend to be veridical. If one holds that (at least some) states of agentive self-awareness have directive structure, reliability will be cashed out in terms of success rather than truth. On a directive account, states of agentive self-awareness are reliable to the extent that they tend to bring about the state of affairs they represent. If one favors a raw feels account, the question

of reliability simply does not arise. Phenomenal feels are not about anything and have no satisfaction conditions; they are simply present or absent.

Given that the debate over the reliability (or lack thereof) of states of agentive self-awareness has mostly been conducted within a descriptive framework, in what follows I will keep to that framework. The question of how reliable states of agentive self-awareness are may yield different answers depending on how much we pack into their contents. The more agentive content includes, the more possibilities for error are created. This question may also yield different answers depending on how closely we take the mechanisms involved in the generation of agentive self-awareness to be related to the mechanisms of action production.

In the sections 2 and 3, we discussed evidence that agentive experience exploits performance-related sensorimotor cues. These would be highly reliable cues to self-agency insofar as they have very direct links to action production (efferent signals) and exploit proprioceptive information that is immune to error through misidentification (proprioceptive information can only be information about oneself). If agentive experiences were based only on efferent cum proprioceptive information they would be very reliable indeed, as it is hard to see how one could experience oneself as actively moving when one isn't. Yet, as we also discussed, these cues contribute rather thin agentive content mostly concerned with active bodily movements, the body parts involved and the timing of their movements.

Perceptual cues largely contribute to enriching the contents of agentive experience. We do not just experience ourselves as moving but also as producing effects in the world. Matches between the predicted perceptual effects of our movements and perceptual events in the world may thus play a central role in agentive experience. Yet, perceptual events resulting from one's own and others' actions are not qualitatively different and keeping them apart is not trivial. Reliance on perceptual cues therefore opens the possibility of errors that can be either

false positives as when have an experience of agency for events actually produced by someone else's action or false negatives as when we lack an experience of agency for an event we actually brought about. The latter kind of error can also happen because we sometimes fail to correctly predict the perceptual consequences of our action. If I press a switch expecting the light to go on and it doesn't, I may fail to notice that by pressing the switch I turned the ventilator on and thus lack a sense of agency for that event.

We can act and lack an experience of agency for perceptual effects that our action brought about or have an experience of agency for perceptual effects that were actually produced by someone else's action. But can we have a sense of agency when we do not act at all? Experimental work suggests that indeed we can. For instance, Wegner and colleagues (Wegner et al., 2004) devised a " helping hands" experiment, where participants watched themselves in a mirror while another person behind them, hidden from view, extended hands forward on each side where participants' hands would normally appear. The hands performed a series of movements. When participants could hear instructions previewing each movement, they reported an enhanced sense of agency for the action compared to those cases where the instructions did not match the hand movement or followed it. These results suggest that when sensorimotor cues and perceptual cues are in conflict, the latter might dominate and trump the former.

However, reliability does not require infallibility and in normal environments such errors may not be systematic enough as to make agentive experiences generally untrustworthy. Those who contend that the sense of agency is deeply flawed tend to attribute richer content to states of agentive self-awareness. Thus, Wegner's claim that the conscious will is an illusion comes with a high-loaded view of what agentive content includes. As several commentators have noted (Bayne, 2006; Carruthers, 2007), he seems to have a number of different illusions in

mind. The experience of freedom and the experience of conscious mental causation appear to be his main targets.

To evaluate claims that these are illusory, one must consider several issues. What are exactly these contentious contents? Are they really the contents of agentive experiences or rather the contents of agentive judgments that interpret agentive experiences? If the latter, are they part and parcel of the folk-psychological conception of agency and thus presumably widely shared or rather the products of philosophical interpretations of folk-psychological notions? Finally, what is the empirical evidence relevant to their assessment?

Free agency

Let us start with the experience of freedom. Here, Wegner's contention seems to be that we experience ourselves as acting freely, but that for this experience to be veridical we would need metaphysically free will, understood as an uncaused cause of action. Libet's experiments (Libet, 1985; Libet, Gleason, Wright, & Pearl, 1983), however, showed that conscious intentions to act (W-judgments) are first experienced on average 200 milliseconds before movement onset but are reliably preceded by several hundred milliseconds by a negative brain potential, the so-called ‘readiness potential’. Either the readiness potential causes the action directly and causes the conscious experience of willing as a mere epiphenomenal accompaniment, or it causes the conscious experience which in turn causes the action. But either way, the conscious will doesn’t initiate the action and thus the action isn't performed freely. Recent experiments (Soon, Brass, Heinze, & Haynes, 2008a, 2008b) also showed that although subjects who had to decide between two actions reported having made a conscious decision on average 1000 ms before action onset, the outcome of their decision was encoded in brain activity of prefrontal and parietal cortex up to 10 s before it entered awareness. Following the same line of reasoning, one could say once again that since conscious decision

follows specific unconscious brain activity rather than precede it, it cannot be free. Therefore, according to Wegner, the experience of freedom is illusory.

Let us note first that it is very doubtful that the experience of freedom, if there is such a thing, includes such a metaphysically loaded notion of freedom. It may well be that what we call the experience of freedom is something much more modest. One may attempt a negative characterization of the experience of freedom and suggest that to experience oneself as acting freely is *not* to experience oneself as compelled to act, in the way that people suffering from obsessive compulsive disorder (OCD) seem to be experiencing their compulsive actions. Indeed, if we take the phenomenology of compulsive actions to be the negative image of the phenomenology of freedom, what seems central is not the felt causal origin of the action but its uncontrollability or unstoppability. People with OCD acknowledge their compulsive actions as originating within their own mind (and not as imposed by outside persons as in delusions of control), but they experience them as uncontrollable – however hard they try to resist performing them, they fail in their attempts. If one's experience of freedom is at bottom the experience that one has control over the actions one performs, can stop an action one is about to engage in, or correct or abort it if need be an ongoing action, the experience of freedom does not appear to qualify as a systematic illusion. Indeed, even Libet acknowledges that we can veto actions and, apart from ballistic actions that cannot be stopped or corrected once started, we are normally able to exert control over our ongoing actions.

But here we are a far cry from the metaphysically loaded notion of freedom that is under attack. We might not err so much in our experiences of freedom than in the interpretations of these experiences that transpire in our agentive judgments. Perhaps then, our folk notion of freedom is incompatibilist and we spontaneously interpret the actions we experience as free as caused by conscious decisions or intentions that are themselves uncaused. This, of course, presupposes that we are natural incompatibilists. But is it really the case? Philosophers

working in the field of experimental philosophy have begun using methods borrowed from psychology to probe ordinary intuitions concerning freedom in a controlled and systematic way. But results so far have not been clear-cut (Feltz, Cokely, & Nadelhoffer, 2009) with some studies suggesting that we are natural compatibilists (Nahmias, Morris, Nadelhoffer, & Turner, 2005; Nahmias, Morris, Nadelhoffer, & Turner, 2006) and others that we are natural incompatibilists (Nichols & Knobe, 2007). From reading this literature, one gets the impression that what people really care about is moral responsibility, for which they take freedom to be necessary, and that they are willing to embrace either compatibilism or incompatibilism as long their doing so helps preserve moral responsibility.

If rather than hardcore incompatibilists or compatibilists, people are merely metaphysical opportunists with respect to freedom, their agentive judgments that they are acting freely need not be metaphysically loaded one way or the other and therefore need not be systematically mistaken.

Conscious mental causation

Another target of Wegner's claim that the conscious will is an illusion is the idea that we have an experience of conscious mental causation. While it is doubtful that a specific metaphysical conception of freedom is part and parcel of our folk-psychology, it is beyond doubt that the idea of conscious mental causation is a core element of our folk-psychology: we firmly believe that our own distinctly mental properties are causally efficacious in the production of our behavior. As pointed out by Hohwy (2004), however, the conception of mental causation that people are so firmly attached to often seems rather poorly articulated. Hohwy usefully distinguishes two sets of aspects in our conception of mental causation. Elements in the first set constitute what he calls the narrow conception of mental causation: we believe in mental causation because we successfully explain and predict behavior in the mental state terms of

common-sense folk-psychology and this success is best explained if these mental states are causally efficacious (reasons as causes), that is if events can be causes in virtue of being instantiations of mental properties. These features of the narrow conception of mental causation do not fully explain the attraction that the idea of mental causation exerts on us. To understand this attraction one must consider a broader conception of mental causation that that appeals to phenomenal features.

According to Hohwy (2004), we further believe that our beliefs and desires and their contents are current triggering causes of our behavior (*here and now causation*), that we are often aware of the mental property that was causally efficacious in causing bodily behavior (*awareness of mental causation*), that we voluntarily select, or endorse, which mental properties are going to cause behavior and control how they do it (*voluntary selection/endorsement and control of mental causes)*; and finally that we generally reliably judge that there is mental causation when there is mental causation (*reliable tracking of track mental causation*). In other words, we do not just believe in mental causation, we believe in conscious mental causation.

It is this latter set of features of our conception of mental causation that appear to be the immediate target of Wegner's skepticism. But are our beliefs in the existence of these features themselves grounded in our agentive experience? Do our agentive experiences include an experience of mental causation that is simply endorsed by these beliefs? Are the beliefs best seen as interpretations of agentive experiences in the light of our folk-psychological conception of mental causation, committing us to more than what our basic agentive experiences actually contain? Opinions are divided. Some theorists, including Hohwy (2004), seem happy to countenance experiences of mental causation, i.e. experiences of one's actions as caused by one's conscious mental states. Others (Horgan, Tienson, & Graham, 2003; Wakefield & Dreyfus, 1991) strongly deny experiencing their actions as caused by their

mental states. Instead, they claim that they experience their actions as caused by themselves or as having their sources in themselves. We may well judge that our actions are caused by our mental states, but such judgments would be more than simply endorsements of agentive experiences. If we allow that we enjoy experiences of mental causation and Wegner's criticisms are warranted, then our agentive experience itself is deeply flawed. If we deny the existence of experiences of mental causation, our agentive judgments regarding mental causation may be shown to systematically in error, without this affecting the probity of our agentive experiences.

Among Wegner's arguments, the argument from automaticity is perhaps the most successful in casting doubt on conscious mental causation. Automaticity is pervasive in our life. In many cases, cognitive goals are triggered by environmental features and our behavior is controlled by automatic mental processes that bypass consciousness (Bargh & Chartrand, 1999; Bargh & Ferguson, 2000). Furthermore, automatic activation of goals creates a state that is functionally very much like conscious activation of goals; motivating behavior and higher mental processes involved in goal-directed behavior, such as maintaining the goal active and monitoring its progress (Bargh & Chartrand, 1999; Custers & Aarts, 2005).[4] If automaticity is pervasive, subjects will often misidentify the reasons they acted as they did, or fail to identify any reasons at all, because those reasons are activated without being consciously accessed.

As pointed out by Hohwy (2004), automaticity poses no threat to the idea of mental causation, insofar as activated goals are mental representations that, when activated, produce their effect qua being mental properties. Automaticity, however, poses a threat to conscious mental causation and to the broad conception of mental causation according to which we are

[4] Recent studies (Aarts, Custers, & Marien, 2009) also provide evidence that mechanisms underlying nonconscious goal pursuit promote experiences of self-agency through the same kind of matching processes that conscious goal pursuit does.

generally aware of the mental states that cause our actions and exert voluntary control on mental causation. How serious the threat depends on how pervasive automaticity is. It is certainly fairly widespread but could it really be the case that all mental causation is automatic? Wegner is certainly right that we overestimate the role conscious mental causation plays in our life. But are we also hopelessly wrong in thinking that conscious mental causation even exists? Indeed, we could turn the argument from automaticity on its head and argue that automaticity is what makes conscious mental causation possible. Conscious processes are known to be slow, resource-demanding, and to have limited capacity compared to automatic processes. We would accomplish very little if we had to rely entirely on conscious processing. Instead, automatic processes make us free to apply our precious conscious resources to issues that matter to us. Perhaps indeed, one reason we overestimate the role played by conscious mental causation generally is that conscious mental causation is at play in matters we really care about.

5. Conclusion

On our folk-psychological conception of ourselves as agents, we are conscious agents both in the sense that our actions are caused by conscious mental states that we voluntarily select and control and in the sense that we have a conscious experience of agency for our actions. Indeed, on this picture conscious agency as conscious mental causation of action and conscious agency as conscious experience of ourselves as agents are really two sides of the same coin: conscious agency in the latter sense is taken to be an experience of oneself as a conscious agent in the former sense.

Empirical work casts a heavy shadow on this portrait of the self as an agent. On the more radical interpretation of these empirical findings, there is simply no such thing as conscious mental causation. Our self-portrait is in effect a vanity picture and our experience of self-

agency a systematic illusion, the result of an elaborate hoax our mind plays on us. More plausibly, I think, empirical findings demand that we seriously retouch this self-portrait, not that we shatter it altogether. On the one hand, they show that conscious mental agency, understood as conscious mental causation, is neither the unique nor the most common form human agency takes. On the other hand, they also suggest that the folk-psychological depiction of the conscious experience of agency as the experience of conscious agency is misleading. In many cases, the experience of agency is something more basic. We may experience an action as ours, without *experiencing* it as caused by conscious mental states and done for this or that reason, although we can readily *interpret* it in that way.

With this touched up portrait, we may also see the link between agency and consciousness of agency in a new light. On the folk-psychological picture of agency as conscious mental causation, the link is constitutive. Indeed, on this picture it is unclear why there should be conscious experiences of agency over and above agency. If agency is by definition conscious, what is the need for an extra-layer of consciousness – a *conscious* experience of *conscious* agency? On the radical reinterpretation of agency Wegner sometimes seems to advocate, we are also left to wonder why we should bother weaving a fictitious conscious story, when actual agency operates quite independently of consciousness. But if human agency is characterized by the interplay of automatic and conscious processes, experiences of agency may play an important role in their integration. If we lacked altogether even a minimal sense of agency for automatically triggered actions, our agentive selves would be but fragmented islands on a sea of automaticity.

References

Aarts, H., Custers, R., & Wegner, D. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition, 14*(3), 439-458.

Bargh, J., & Chartrand, T. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462-479.

Bargh, J., & Ferguson, M. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126*(6), 925-945.

Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will. In S. Pockett & W. Banks & S. Gallagher (Eds.), *Does Consciousness Cause Behavior?* (pp. 169-185). Cambridge, MA: MIT Press.

Bayne, T. (2008). The phenomenology of agency. *Philosophy Compass, 3*(1), 182-202.

Bayne, T. (2009). The Sense of Agency. In F. Macpherson (Ed.), *The Senses*. Oxford: Oxford University Press.

Bayne, T., & Pacherie, E. (2007). Narrators and comparators: The architecture of agentive self-awareness. *Synthese, 159*(3), 475-491.

Blakemore, S., & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology, 13*(2), 219-224.

Blakemore, S., Wolpert, D., & Frith, C. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience, 1*, 635-640.

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport, 11*(11), R11-R16.

Bruner, J., & Goodman, C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*(1), 33-44.

Carruthers, P. (2007). The illusion of conscious will. *Synthese, 159*(2), 197-213.

Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89*(2), 129-142.

Davidson, D. (1980). *Essays on actions and events*: Oxford University Press, USA.

De Vignemont, F., Tsakiris, M., & Haggard, P. (2006). Body mereology. In G. Knoblich & I. Thorton & M. Grosjean & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 147-170). New-York: Oxford University Press.

Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel & P. Cole & D. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103-115). Hillsdale, NJ: Erlbaum.

Feinberg, T. E., Roane, D. M., & Ali, J. (2000). Illusory limb movements in anosognosia for hemiplegia. *Journal of Neurology Neurosurgery and Psychiatry, 68*(4), 511-513.

Feltz, A., Cokely, E., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language, 24*(1), 1-23.

Fourneret, P., de Vignemont, F., Franck, N., Slachevsky, A., Dubois, B., & Jeannerod, M. (2002). Perception of self-generated action in schizophrenia. *Cognitive neuropsychiatry, 7*(2), 139-156.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia, 36*(11), 1133-1140.

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience, 11*(11), 3656-3666.

Frith, C. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition, 14*(4), 752-770.

Frith, C., Blakemore, S., & Wolpert, D. (2000a). Abnormalities in the awareness and control of action. *Philos Trans R Soc Lond B Biol Sci, 355*(1404), 1771-1788.

Frith, C., Blakemore, S., & Wolpert, D. (2000b). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews, 31*(2-3), 357-363.

Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass, 2*(2), 347-357.

Gazzaniga, M., & LeDoux, J. (1978). *The integrated mind*: Plenum Publishing Corporation.

Haggard, P., & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and Cognition, 12*(4), 695-707.
Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience, 5*(4), 382-385.
Hohwy, J. (2004). The experience of mental causation. *Behavior and Philosophy, 32*(2), 377.
Horgan, T., Tienson, J., & Graham, G. (2003). The Phenomenology of First-Person Agency. In S. Walter & H.-D. Heckmann (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action.* (pp. 323-340). Exeter Imprint Academic.
Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. *Experimental Brain Research, 192*(3), 527-532.
Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT press
Knoblich, G., & Kircher, T. T. J. (2004). Deceiving oneself about being in control: Conscious detection of changes in visuomotor coupling. *Journal of Experimental Psychology-Human Perception and Performance, 30*(4), 657-666.
Knoblich, G., & Repp, B. H. (2009). Inferring agency from sound. *Cognition, 111*(2), 248-262.
Knoblich, G., Stottmeister, F., & Kircher, T. (2004). Self-monitoring in patients with schizophrenia. *Psychological Medicine, 34*(08), 1561-1569.
Leube, D., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. (2003). The neural correlates of perceiving one's own movements. *Neuroimage, 20*(4), 2084-2090.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action *Behavioral and Brain Sciences, 8*(4), 529-539.
Libet, B., Gleason, C., Wright, E., & Pearl, D. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain, 106*(3), 623.
Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks, 9*(8), 1265-1279.
Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives, IX*, 185-200.
Moll, A. (1889). *Hypnotism*. London: Walter Scott.
Mossel, B. (2005). Action, control and sensations of acting. *Philosophical Studies, 124*(2), 129-180.
Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology, 18*(5), 561-584.
Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research 73* 28 - 53.
Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous, 41*(4), 663-685.
Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179-217.
Pacherie, E., Green, M., & Bayne, T. (2006). Phenomenology and delusions: Who put the 'alien'in alien control? *Consciousness and Cognition, 15*(3), 566-577.
Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology, 16*, 337-348.
Pockett, S., Banks, W. P., & Gallagher, S. (Eds.). (2006). *Does consciousness cause behavior?* Cambridge, Mass.: MIT Press.
Proust, J. (2003). Perceiving intentions. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness* (pp. 296-320). Oxford: Oxford University Press.
Roessler, J., & Eilan, N. (Eds.). (2003). *Agency and self-awareness: issues in philosophy and psychology*: Clarendon Press.
Roser, M., & Gazzaniga, M. (2004). Automatic brains-interpretive minds. *Current Directions in Psychological Science, 13*(2), 56-59.
Roser, M., & Gazzaniga, M. (2006). The interpreter in human psychology. *The Evolution of Primate Nervous Systems. Oxford; Elsevier*.
Sass, L. (1992). *Madness and modernism: Insanity in the light of modern art, literature, and thought. .* New York: Basic Books
Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition, 110*(1), 74-83.

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition, 94*(3), 241-255.
Searle, J. (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge: Cambridge University Press.
Sebanz, N., & Prinz, W. (Eds.). (2006). *Disorders of volition*. Cambridge, MA: MIT press.
Siegel, S. E. (2007). Symposium on the Phenomenology of Agency. *Psyche, 13*(1).
Slachevsky, A., Pillon, B., Fourneret, P., Renie, L., Levy, R., Jeannerod, M., & Dubois, B. (2003). The prefrontal cortex and conscious monitoring of action - An experimental study. *Neuropsychologia, 41*(6), 655-665.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008a). Unconscious determinants of free decisions in the human brain. *International Journal of Psychology, 43*(3-4), 238-238.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008b). Unconscious determinants of free decisions in the human brain. *Nature neuroscience, 11*(5), 543-545.
Stephens, G., & Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. Cambridge, MA: MIT Press.
Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219-239.
Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology-Human Perception and Performance, 31*(1), 80-91.
Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition, 96*(3), 215-231.
Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition, 15*(2), 423-432.
van den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition, 85*(2), 177-187.
Wakefield, J., & Dreyfus, H. (1991). Intentionality and the phenomenology of action. In E. LePore & R. Van Gulick (Eds.), *John Searle and his critics. Cambridge, MA: Blackwell* (pp. 259-270).
Wegner, D. (2002). *The illusion of conscious will*: The MIT Press.
Wegner, D. (2005). Who is the controller of controlled processes. In R. R. Hassin & J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious* (pp. 19–36). Oxford: Oxford University Press.
Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86*(6), 838-848.
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation - Sources of the experience of will. *American Psychologist, 54*(7), 480-492.
Weiller, C., Juptner, M., Fellows, S., Rijntjes, M., Leonhardt, G., Kiebel, S., Muller, S., Diener, H. C., & Thilmann, A. F. (1996). Brain representation of active and passive movements. *Neuroimage, 4*(2), 105-110.