Chapter 14
What is the Human Sense of Agency, and is it Metacognitive?

Valerian Chambon, Elisa Filevich and Patrick Haggard

Abstract Agency refers to an individual's capacity to initiate and perform actions, and thus to bring about change, both in their own state, and in the state of the outside world. The importance of agency in human life cannot be understated. Social responsibility is built on the principle that there are "facts" of agency, on which individuals can generally agree. At the individual level, the experience of agency is considered a crucial part of normal mental life. Abnormal sense of agency (SoA)—such as in the well-documented "delusion of control"—is recognised as one of the key symptoms of mental disorders. Yet, beyond abnormalities of control that pertain to psychiatric conditions, normal SoA can be easily fooled. Errors in agency attribution and agency experience have received much attention in recent experimental literature. In everyday life, coincidental conjunctions between our actions and external events commonly occur. The fact that the SoA can be over or underestimated, or that judgements of agency can be wrong, testifies to a significant gap between what individuals think or believe their control capabilities are, and what these capabilities really are. The ability to experience these computations as the causes driving and shaping our actions may account for our ability to correct our behaviours when, precisely, they seem to escape our control. In this sense, any reliable theory about human agency must explain how we can sometimes be deluded about our own agency, but also must account for why we are not deluded all the time. In this chapter, we first identify which signals may contribute to an SoA, and how they might be integrated. We will ask whether human cognition of agency is best analysed as an *experience* or as an *inference*. We evaluate the existing data in relation to two contrasting accounts

V. Chambon
INSERM, ENS, Paris, France

E. Filevich
Max-Planck Institute for Human Development, Berlin, Germany

P. Haggard (✉)
ICN-UCL, London, UK
e-mail: p.haggard@ucl.ac.uk

S. M. Fleming and C. D. Frith (eds.), *The Cognitive Neuroscience of Metacognition*,
DOI: 10.1007/978-3-642-45190-4_14,

for agency, namely prospective versus purely retrospective approaches. We draw on two major classes of data throughout: psychological data that aims to capture the experience of agency, and physiological data that aims to identify the neural basis of this experience. Finally, we will consider whether the human SoA should really be called 'metacognitive'. In particular, we directly compare key features of metacognition of agency with perceptual metacognition.

14.1 What is Agency?

Agency refers to an individual's capacity to initiate and perform actions, and thus to bring about change, both in their own state, and in the state of the outside world. The importance of agency in human life cannot be understated. Societies depend on the idea that there are "facts" of agency, on which individuals can generally agree. This allows societies to hold individuals responsible for their own actions and for their consequences, thus rewarding or punishing the individual for what they do. Legal responsibility, and payment for labour provide two pervasive examples.

There are at least two aspects of agency. First, agency is an objective fact, demonstrated by individuals' behaviours and the consequences of those behaviours. But agency has a first-person component as well: it involves distinct cognitive processes and subjective experience unique to the agent. The experience of agency is considered a crucial part of normal mental life. Abnormal sense of agency (SoA)—as in the well-documented "delusion of control"—is recognised as one of the key symptoms of mental disorders. Further, links between SoA and health and well-being in the general population have been clearly established [7]. Nevertheless, the basis of the SoA is poorly understood.

Here we investigate what aspects of agency, if any, are metacognitive. We do this by analysing agency into a number of components, and by investigating how each component is computed in the human brain. We use two major distinctions to investigate the basis of SoA. First, we distinguish the types of *signals* contributing to SoA. This allows us to distinguish prospective SoA based on predictive signals linked to action intentions, from retrospective SoA based on action outcomes. Second, we distinguish the types of cognitive *processes* operating on those signals, to ask whether agency is best analysed as an *experience* or as an *inference*. In each case, we ask whether the particular component of agency can be considered metacognitive or not, and why. Finally, we compare the metacognition of agency with the features of the more widely studied perceptual metacognition.

We will draw on two major classes of data throughout: psychological data that aims to capture the experience of agency, and neural data that aims to identify the neural basis of this experience. The ability to *experience* these computations as the causes driving and shaping our actions, may account for the ability to correct our actions when action control is suboptimal. In this sense, any reliable theory about

human agency must also explain how we can sometimes be deluded about our own agency, but also must account for why we are not deluded *all the time*.

To investigate whether SoA is or is not metacognitive, we need a clear definition of metacognition. Under a wide definition, metacognition is the general ability to monitor mental states and processes. This monitoring may be explicit or not. Explicit monitoring leads to meta-representations that allow reflecting upon, commenting about and reporting on the metal processes. Experience monitoring, however, may be implicit and may not allow explicit judgements based on first-order processes. Rather, the operation of the first-order processes is experienced. The concepts of first and second-order processing are central in current work on metacognition. Two main approaches have been taken, towards the study of metacognition. First, in psychophysical tests, experimenters may ask human volunteers to make simple (typically visual) perceptual judgements, and to also report their confidence on each of their responses [32]. Confidence judgements are thought to depend on purely internal aspects of the processing of first-order perceptual input signals. A second important body of work has investigated the relation between knowledge, and "knowing that you know" [42, 43]. In both cases, second-order processing within the brain itself generates an experience that can play a functional role in the organism's mental life and behaviour. In both cases, the distinguishing feature of metacognition is the presence of an internal, first-order signal as the content of a second-order representation or process.

Based on this view, we can now consider (a) which signals contribute to agency, (b) whether SoA is metacognitive in virtue of the nature of those signals, and why, (c) whether the SoA is similar, or essentially different, from other metacognitions, given that its 0th-order contents (i.e. actions and outcomes) differ from the 0th-order contents studied in other well-established areas of metacognition, such as perception (e.g. visual input) and knowledge (e.g. facts about the world). Addressing these questions must inevitably begin with a clear, analytical understanding of SoA.

14.2 Experiences of Agency

Agency can be defined from the point of view of an external observer, as it is related to the objective fact that individuals can make actions, and change their environment. However, agency also involves distinct cognitions and experiences on the part of the agent. Following Synofzik et al. [69] we use the term 'sense of agency' (SoA) to refer to the feeling or experience that individuals may have in relation to their own actions, and to the consequences of their actions, when *they* control those actions. We use the term 'judgement of agency' (JoA) to refer to an explicit judgement made by an individual regarding whether they, or another individual, brought about the action, or the external event. Note that both SoA and JoA are cognitive constructs rather than external physical facts. To this extent they

are both subjective, rather than objective, and both can be wrong. For example, an individual can have an illusion of agency when they in fact are not the agent, as we will see later. Note also that JoA and SoA are normally related. In particular, an individual may judge that they are the agent of an event because they have an SoA with respect to that event. Likewise, and individual may judge they are not the agent, because they lack an SoA. We return to the relation between SoA and JoA later in this chapter.

The relationship between SoA and JoA is also important for the organization of societies. All known human societies depend on attribution of blame, and thus on individual responsibility for action. That is, societies depend on the idea that there are facts of agency, on which individuals can generally agree. Third-person judgements of agency must then have clear and objective truth conditions. However, agreement about judgements of agency and responsibility is only possible if individuals' brains support a subjective experience of agency. Only then will individuals feel and understand their actions and responsibilities, and accept society's third-person judgements of agency. To be useful, judgements of agency must align both with facts of agency, and with SoA, in most cases, though not necessarily in all. Therefore, the experience of agency is considered a crucial part of normal mental life. Abnormal SoA is recognised as one of the key symptoms of mental disorders, and links between SoA and health and well-being have been clearly established [7].

Despite this importance, SoA has only recently been addressed within cognitive science, perhaps because appropriate methods of measurement have been lacking. In particular, the SoA, as in general the experience of voluntary action, has been described as 'thin' and 'elusive'. Few psychophysical studies have sought to identify the factors that influence SoA and JoA.

14.3 Analytic Structure of Agency

14.3.1 Agency Impressionism

As a first step in an experimental analysis of SoA, we should characterise the experience of agency itself, and consider how it can be measured. On one view, agency is an atomic experience, and without any internal analytic structure. It is an impression that individuals directly and authoritatively have in cases where they are in fact responsible for an external sensory event. We call this view *agency impressionism*, by analogy with Michotte's concept of a causal impression [53]. Indeed, it was classically suggested that a direct impression of one's own motoric agency formed the basis for cognition of general causation in the external world [16]. A principal difficulty for agency impressionism is to explain why, if agency is directly perceived, illusions and misperceptions of agency may nevertheless occur [76].

14.3.2 Relational View

An alternative view is based on the *relational* aspect of agency. The facts of agency depend on a particular relation between an individual and an event, expressed by the proposition "I did that". On the relational view, these two components "I" and "that" remain present in the experience of agency itself. SoA is not, therefore, an immediate perceptual experience in the same way that a sound or a smell may be, because it involves a second-level relation between two primary elements, the agent and the event. Relational theories would view SoA as more than an atomic percept. Following Hume's view of causation [39], relations cannot be perceived directly. Rather, the relational view suggests that the mind supplies the relation between agent and event, based on the conjunction of the percepts of the cause (one's own intentional action), and the effect (the action outcome).

The relational view has two strong merits. First, it explains how SoA can be generalised, substituted or extrapolated from one case to another. An individual's SoA when they switch on a light has much in common with their SoA when they switch on a radio, or cause some similar event in the outside world. The "that" in "I did that" can be substituted. The agent and action remain the same, though the outcomes differ, and the feeling of being in control also remains broadly the same. By the same token, the SoA may be similar when an individual uses their hand or their head to switch on the light [34]. Similarly, the "I" in "I did that" can be substituted. One individual's SoA is assumed to be much like another's, so that one can understand another's SoA by observing the relation between their actions and subsequent outcomes [25]. The idea of agency as a relation implies that the key aspects of SoA should remain constant even when the basic content of action and outcome vary.

Recent experiments broadly confirm the relational view. The effects of time delays between action and outcome have been particularly extensively studied. In particular the intentional binding effect [35] reliably shows that actions are perceived as shifted in time towards the outcomes that they cause, while outcomes are perceived as shifted back in time towards the actions that cause them. This temporal attraction emphasises the temporal contiguity and conjunction between action and effect [39]. The effect is reduced or absent in cases of involuntary or passive movement [35]. Equally, when participants make numerical judgements about the interval between action and effect, their judgements show a perceptual compression, relative to intervals that begin with equivalent passive movements [20]. These data provide strong evidence that the relation between action and outcome is indeed a core component of SoA. They are also consistent with a broadly Humean associationist account of SoA. There may be no direct experience of agency over and above the experiences of the action and outcome itself, yet the mind may associate experiences of actions and outcomes so that they stand in a characteristic relation to each other. In the next section, we consider the signals that are related, and how the relation might be computed.

14.3.3 Signals for Agency

A person who grasps the relation "I did that" must be sensitive to two different signals, corresponding to the "I" and the "that". This suggests that SoA presupposes a relation between two quite distinct components, which we call *attribution* and *instrumentality*. Attribution concerns the "I" component. From a signal-processing point of view, someone who grasps that "I did that" must be capable of discriminating between themselves and other agents. That is, they must be able to attribute the outcome to "I", rather than "you", or any other cause. This requires a signal sensitive to *one's own* agency, i.e. some neural event that is present when one is the agent, and only when one is the agent. In philosophy, the direct, first-person access to one's own intentional states provides this signal [61]. In contrast, in neuroscience, awareness of one's own intentions remains controversial [29, 45], and the idea of immunity from error through self-identification has been questioned. Nevertheless, experimental studies show that self-recognition through active movement is superior to that with passive movement [71]. Efferent signals—i.e. signals that are sent from the brain's motor centres via the spinal cord to the muscles—therefore play an important role in discriminating "I" from other agents, and may provide the basis for attribution aspects of agency. Importantly, errors in agency attribution should then occur when efferent signals provide little discriminative information about agency, for example in situations where several people act at once.

Instrumentality refers to the "that" component of "I did that". To have an SoA, an individual must discriminate between events that she did cause, and events that she did not. Again, signals regarding one's intentional actions may play a key role. To know that "I did p", but "I did not do q", it may be sufficient to have access to an efferent signal that correlates well with p, and correlates poorly with q. Several studies have investigated the role of motor identity (i.e. response—stimulus associations), temporal relations and statistical contingency in the representation of agency [73]. In some cases, one can cause something to happen despite not intending to do so. One may even retrospectively acquire an SoA in such cases, by coming to believe that one *had* intended to do so. However, these are cases where SoA is decoupled from facts of agency. The primary task of a metacognitive account of agency is to deal with how our factual agency is experienced [60].

Interestingly, most previous studies of "agency", focus *either* on attribution ("I"), *or* on instrumentality ("that"), but do not clearly distinguish between the two aspects. This has lead to considerable confusion: often studies of "agency" meet with the reaction "that's not what we mean by agency". We believe that, in many cases, this critique should really be translated as "What does your account of instrumentality imply for attribution?", or "What does your account of attribution imply for instrumentality?".

The crucial link between attribution and instrumentality is that both depend on action signals. But a signal-processing approach clearly shows that these signals can provide information of two different kinds. In computing attribution, efferent signals are used to discriminate between agents, and can support explicit

judgements of agency. In computing instrumentality, efferent signals are used to discriminate between outcomes. Efferent signals provide an SoA relating to some outcomes, but not others. The implications of this distinction for metacognition are discussed in Sect. 14.7.

14.4 Computational Models of Agency

14.4.1 Comparator Models

We previously took "I did that" as the cardinal expression of agency. This shows that agency also implies a specific process of action control (corresponding to "did") that relates these terms. Specifically, agency implies a control mechanism that has goals, and that controls actions to achieve them. This concept was successfully formalised as a comparator model [52, 78]. These models translate intentions into outcomes, by continually monitoring whether action consequences occur, or do not occur, as predicted. Though originally formulated as models of motor control, comparator models have also been increasingly used to explain the subjective SoA.

Because of their importance in the agency literature, we will present these models in some detail (see Fig. 14.1). A typical framework comprises two internal models: an inverse model and a forward model (see [30]). The desired goal is first fed to the inverse model which selects appropriate motor commands for achieving the goal. These commands are then executed, by sending them from the brain to the musculature. At the same time, a copy of the motor commands ("efference copy") is fed to the forward model, which predicts their effects. Thus, the motor control system can predict the current state of the body in advance of delayed sensory feedback about the effects of a motor command. This predictive information can then be used in two critical comparisons.

First, it can be compared to the desired goal state, to assess whether further motor commands are required, or whether the action has achieved its goal. Second, it can be compared to sensory information about actual effects of action. This second comparison assesses whether sensory information is or is not a predicted consequence of the current motor command. Crucially, this second comparison can distinguish self-caused sensory events (reafferences) from external events (exafferences). Thus, it functions as an agency-detector: no error means "I did that", while any error signal means "I didn't do that". On this view, agency can be attributed by low-level, pre-reflective mechanisms that learn to predict consequences of motor commands [9, 69].

Several studies confirm a role for motor prediction in agency judgement (see [14] for a review). Introducing a temporal or a spatial transformation between an action and its visual consequences reduces participants' sense of control in proportion to the mismatch induced. In one task, participants received distorted visual

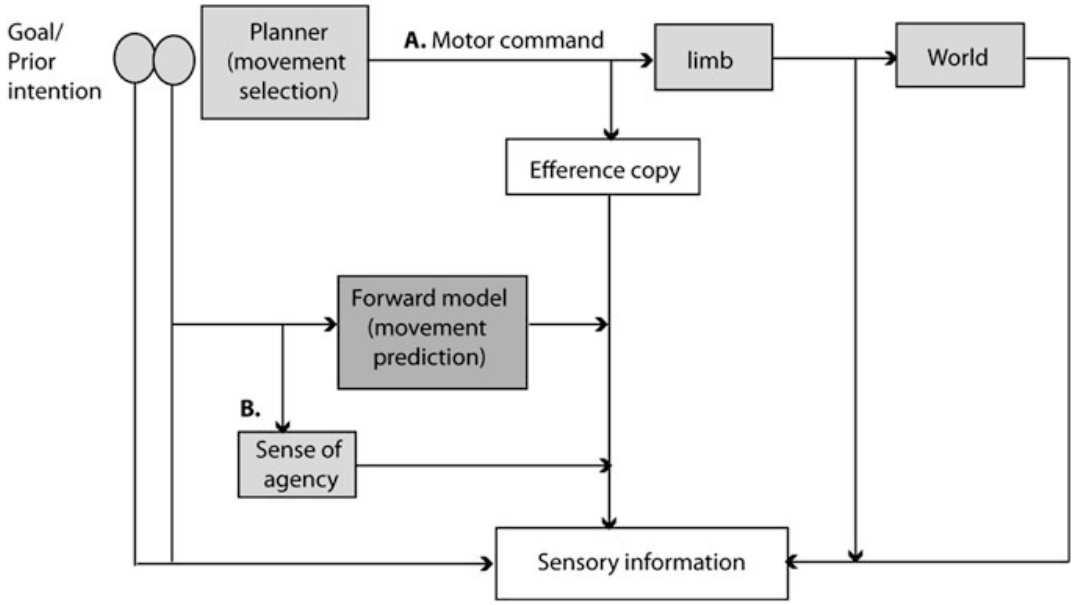

Fig. 14.1 A computational framework for action. Point *A* marks a point after movement selection where conscious awareness of intention might arise. Point *B* marks an integration of efference, predicted feedback and sensory information, which might lead to the sense of agency (adapted from Haggard 2005)

feedback of their hand moving a joystick. When the movement of the virtual hand did not correspond to the subjects' movement [23], or when an angular bias was introduced between the subject's and the virtual hand's movement, participants more readily attributed it to another agent [13, 21, 27, 68]. Note that manipulating temporal relations between actions and outcomes had similar effects [13, 15, 22, 28, 44, 49].

On comparator accounts, a positive SoA is the default operation when no error occurs. It is the experiential output of subpersonal processes that mostly run outside consciousness [69]. Crucially, although SoA relies on real-time motor signals, it can only be *computed* after those signals are compared with reafferent (visual, motor, or proprioceptive) feedback. Thus, a reliable, explicit SoA may only be formed when reafferent signals become available for matching with intentions. Thus, one cannot feel agency over any event until that event has been registered and processed in the brain. Although agency is informed by online signals about motor guidance and control, it can only be *retrospectively* attributed [9].

14.4.2 "Belief-like" Models

An alternative model treats agency not as a result of sensorimotor computations, but as an inference about authorship. Prior thought about an event, and general predictability of the event boost the experience of agency [3, 48, 63, 75]. This

series of findings strongly suggests that agency does not simply depend on predictive motor signals. Instead, agency may be based on a general mechanism for estimating event likelihoods. When prior conscious thought about doing X co-occurs with X itself, a causal relationship is retrospectively *assumed* [39]—between the self and an external event, so that the event is inferred as having been caused through one's own will or action [74]. On this view, the experience of action would be necessarily *reconstructed* as an output of this secondary, belief-fixation mechanism. Thus, both belief and comparator models are reconstructive. Agency attribution, as a way of rationalising our actions and experiences, could thus primarily depend on conceptual, reflective processes or states—such as ad hoc theorising about oneself [72] or personal background beliefs [31]—, and not only on a signals within comparator. Importantly, belief-based models of agency allow that SoA is a consequence of JoA, rather than a cause.

14.5 Neural Bases of Agency

Reduced SoA following spatial and temporal mismatches between anticipated and actual action consequences is associated with increased activation in the angular gyrus (AG, [21–23]. Activation of AG should code for feelings of non-agency under ambiguous experience, rather than for positive self-agency experience [56]. The cerebellum may also signal discrepancies between predicted and actual sensory consequences of movements [5, 6, 59]. Other candidates for the comparator role have also been suggested, including premotor cortex [18, 19]. Interestingly, the opposite pattern of activation has been observed in the insula. Insula activation is positively correlated with control felt by subjects over visual consequences of their action [21]. However, this activation has also been interpreted as related to sense of body ownership, rather than agency [70].

By contrast, the belief-like account of agency might recruit higher cortical centres such as the prefrontal cortex (PFC), which provide conscious monitoring [67] rather than sensorimotor integration. Specifically, the dorsal lateral part of the PFC has been implicated in conflict monitoring and detection such as between intention and sensory outcome (e.g. [24, 66]). The supplementary and pre-supplementary motor areas might also be recruited when motor intention matches with a sensory feedback, to give rise to the intentional binding, mentioned before [56]. Finally, the interplay between these medial frontal areas and the PFC may be crucial for SoA. On one view, mismatches in cases of non-agency detected by AG are transmitted to PFC where alternative accounts of agency would be computed retrospectively (see [59]).

We may ask whether comparator models and belief models are truly metacognitive. That is, are judgements of agency generated by second-order processes that process purely internal signals? In the case of comparator models, the answer is a clear 'yes': the model is based on an efference copy and an internal predictor that operates in advance of action itself. If a signal that roughly corresponds to an

internal intention contributes to SoA or JoA, then these states are, at least partly, metacognitive. However, it is much harder to prove that any particular JoA crucially depends on these internal signals. In particular, the internal signals are highly correlated with signals provided by the sensorimotor action itself. Thus, alternative, non-metacognitive accounts based on reconstructive inference from non-internal signals are always available. For example, if I judge that I switched on the light, the comparator model would view the judgement as driven by an efferent signal corresponding to the intention to switch on the light. However, the same judgement could also be an inference or assumption that one had switched on the light, driven by one's knowledge that it was dark, one's first-level experience that one's hand is touching the light switch, and one's first-level experience that the lights have come on [74].

Belief-like models need not be metacognitive. As the above example of the light switch shows, the belief model might begin with a "prior conscious thought" that it is dark. Somatosensory feedback from the hand on the switch, and visual feedback from the lights coming on are then sufficient to infer agency. This inference then leads to reconstruction of a first-person, explicit, judgement of agency. This judgement need not be related to first-order internal processes. Thus, previous studies linked to comparator and belief models provide only modest support for the view of agency as a form of metacognition.

In our view, only one class of evidence conclusively demonstrates that SoA is based on the internal efferent signals of the comparator model. Patients with anosognosia for hemiplegia [4] report an SoA over actions which they are, in fact, unable to make because of paralysis. This experience appears to be driven the internal signal corresponding to the intention to act [33]. Because of deficient feedback-based monitoring due to the lesion, this signal is sufficient to generate an SoA in the patients [26].

In addition to evidence from patient populations, we consider an alternative, more recent class of evidence from studies in healthy volunteers in the Sect. 14.6.

14.6 Beyond Comparators: Experiential Metacognition of Agency

14.6.1 Action Selection Contributes to Feelings of Control

Previous studies have shown that judgements of agency tend to be related to how participants think that they perform in a task [51]. Similarly, errors in task performance may lead to a *feeling* of something dysfluent during the task, without any explicit awareness of an error, and without ability to explicitly report the error (see [51]). The term 'epistemic feeling' has been coined to describe this subjective, online, experience of an error [12, 58]. Importantly, these epistemic feelings strongly influence the SoA, as shown by recent subliminal priming studies.

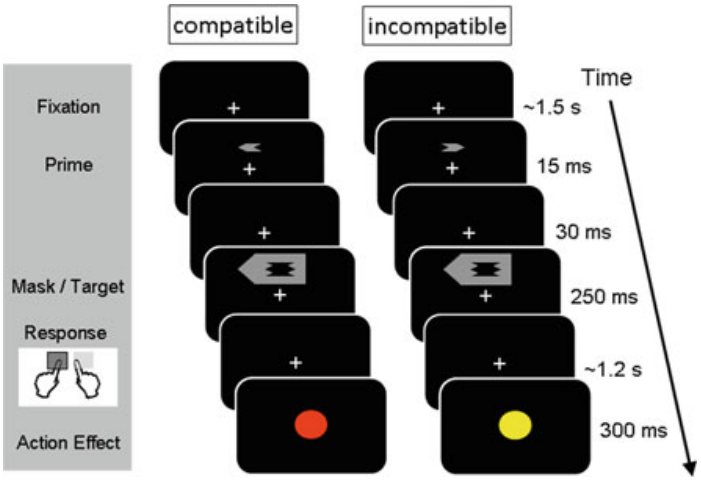

Fig. 14.2 Schematic of trial procedure and stimuli (cued-choice conditions only). Example trials from the two possible combinations of the prime-action compatibility (compatible: *left panel*; incompatible: *right panel*). The appearance of the effect was randomly jittered 150, 300 or 450 ms after the keypress to avoid ceiling effects in perceived control. Adapted from Wenke et al. [77]

We have recently identified a situation where an avowedly internal signal appears to contribute to the SoA [77]. We showed that the SoA could be modulated by using subliminal priming to affect the *fluency* of action selection processes. Interestingly, this procedure allowed us to manipulate the subjective sense of control, without manipulating the *predictability* of action outcomes. We interpret this as an implicit, non-conceptual form of metacognition [9, 58].

In this experiment, participants pressed left or right keys in response to left- or right-pointing arrow targets. Prior to the target, subliminal left or right arrow primes were presented, unbeknownst to the subject. Prime arrow directions were either identical (compatible condition) or opposite (incompatible condition) to the subsequent target (Fig. 14.2). Responding to the target caused the appearance of a colour after a jittered delay. The colour patch can thus be considered as the action outcome. The specific colour shown depended on whether the participant's action was compatible or incompatible with the preceding subliminal prime, but did not depend on the prime identity or the chosen action alternative alone. Unlike previous studies, therefore, the primes did not predict action effects, nor could any specific colour be predicted on the basis of the action chosen. Participants rated how much control they experienced over the different colours at the end of each block [77].

Analyses of reaction times showed that compatible primes facilitated responding whereas incompatible primes interfered with response selection. More importantly, priming also modulated the sense of control over action effects: participants experienced more control over colours that followed actions compatible with the preceding primes than over colours that followed prime-incompatible actions. Thus, subliminal priming made action selection processes more or less *fluent*, and this modulation of fluency affected the sense of control over action outcomes.

These results have several important cognitive implications. First, they suggest that the SoA depends strongly on processes of action selection that necessarily occur *before* action itself. Second, strong SoA may be associated with fluent, uncontested action selection. In contrast, conflict between alternative possible actions, such as that caused by incompatible subliminal priming, may reduce the feeling of control over action outcomes. Third, this prospective contribution of action selection processes to SoA is distinct from predicting the outcomes of action, since action outcomes were equally (un-) predictable for compatible and incompatible primes. That is, these primes did not prime effects of action as in previous studies (e.g. [1, 46, 62, 76]). Therefore, participants could not retrospectively base their control judgements on match between primes and effects alone. Rather, their stronger experience of control when primes were compatible could only be explained by the *fluency* of action selection—i.e. by a signal experienced *before* the action was made, and the effect was displayed.

Finally, participants did not consciously perceive the subliminal primes. Therefore, participants' sense of control could not be based on (conscious) beliefs about the primes. Instead, action priming itself presumably directly influenced the subjective sense of control. Pacherie [61] (see also [69]) has suggested that action selection conflict need not necessarily be conscious [57]. Such conflict may elicit the feeling "that something is wrong", without (necessarily) leading to knowledge about *what* is wrong. Wenke et al.'s study shows that subjects can rely on this first-person, implicit feeling to make judgements about their own control over action effects.

14.6.2 Dissociating Fluency of Action Selection from Performance Monitoring

Monitoring fluency signals generated *during* action selection could therefore be an important marker for the experience of agency. If so, agency would clearly have a metacognitive component, because these signals are generated internally by the process of action selection. However, it is also possible that participants might have estimated agency based on implicit monitoring of their own performance, such as their reaction times (RTs). Since RTs are lower on compatibly primed trials [17, 64, 65], participants would therefore feel more control on compatible trials, because they respond more rapidly. On this second view, agency would depend on *retrospective* monitoring of action execution performance [50], not on *prospective* monitoring of premotor fluency signals. Importantly, SoA would have a metacognitive aspect according to the latter view, but not the former.

To distinguish between these two accounts of sense of control, Chambon and Haggard [10] used an experimental procedure that dissociated fluency of action selection from performance monitoring. Specifically, they increased the interval between mask and target to take advantage of a Negative Compatibility Effect

(NCE) in priming. Longer mask-target latencies *increase* RTs following compatible primes, relative to incompatible primes [65]. By combining this factor with Wenke et al.'s design for assessing sense of control, it was possible to directly compare the contrasting retrospective (performance monitoring) and prospective (action selection) accounts. Specifically, if sense of agency depends on intentional fluency, it should be greater when actions are compatibly versus incompatibly primed, irrespective of whether priming benefits or impairs performance. Alternatively, if SoA depends only on performance monitoring, it should be stronger for rapid versus slower responding, irrespective of whether priming is compatible or incompatible with the action executed.

Crucially, reversing the normal relationship between prime-target compatibility and RTs did not alter subjective sense of control. Thus, in compatible NCE trials, participants experienced *stronger* control despite *slower* response times and higher error rates, compared to incompatible NCE trials. These results suggest that the feeling of control normally experienced by subjects on compatible trials does not depend on retrospectively monitoring performance, thereby strengthening the evidence for a prospective contribution of action selection fluency to SoA.

In both Wenke et al. [77] and Chambon and Haggard [10] experiments, priming did not influence the actual objective level of control that participants had over the colours presented after their actions. Indeed, the contingency between action and colour effect was similar for compatibly primed and incompatibly primed trials. So the prospective sense of control identified in these experiments is in fact an illusion of control, since it is not based on differences in the actual statistical relation between action and effect. In other words, action selection is irrelevant to actual action/effect contingency, and thus to the agent's actual ability to drive external events. However, this prospective sense of control may nevertheless be a convenient proxy for actual control, because we often just know what to do and what will happen next. In that sense, *fluent* action selection is generally a good advance predictor of actual statistical control over the external environment [37]. If prospective agency is a particular conscious experience generated by action programming, which we learn to use as a convenient marker of our own factual agency, it might indeed qualify as a metacognition.

14.6.3 Prospective Agency: Neural Underpinnings

Taken together, these findings suggest that neural activity in action preparation circuits *prospectively* informs agency, independent of outcome predictability of the outcome, and actual performance. Tracking dysfluency in action selection networks [54, 59] could be the basis for this prospective SoA. Recently, Chambon and collaborators [11] adapted the prospective agency paradigm for functional neuroimaging (fMRI). They studied whether the angular gyrus (AG), which has been shown to compute *retrospective* agency by monitoring mismatches between actions and subsequent outcomes [21, 22], may also code for a *prospective* sense

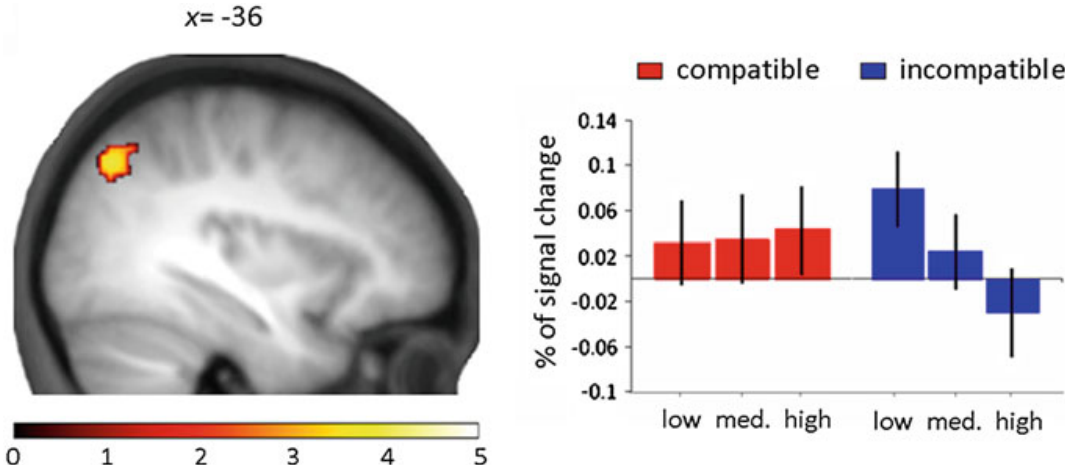

Fig. 14.3 Parametric interaction of control and compatibility in the angular gyrus (AG). Left AG is differentially modulated by participants' control ratings depending on how fluent action selection is; scale shows *t*-value. Adapted from Chambon et al. [11]

of control, by monitoring action selection processes in advance of the action itself, and independently of action outcomes. This would inform one *whether one's actions are appropriately following through one's original intentions*. If a dysfluency, or causal break between intention and action occurs, SoA over outcomes would be reduced.

Again, participants experienced greater control over action effects when the action was compatibly versus incompatibly primed. More importantly, this prospective contribution of action selection processes to SoA was accounted for by exchange of signals between specific frontal action selection areas and the parietal cortex. First, Chambon et al. found that activity in the angular gyrus was sensitive to mismatches, but not matches, between prime arrow and actual response to the target arrow. Moreover, this activity due to the prime-target mismatch predicted the magnitude of subsequent sense of control: for incompatible trials only, activity in the AG decreased as sense of control over outcomes increased (Fig. 14.3a). Importantly, this neural coding of non-agency occurred at the time of action selection *only*, as in Wenke et al.'s original experiment.

Second, activity in the AG (signalling non-agency) in incompatible trials was negatively correlated with activity in the dorso-lateral prefrontal area (DLPFC) (Fig. 14.3b). This pattern of fronto-parietal interaction would reflect contribution of action selection brain areas to sense of control. Indeed, DLPFC has long been associated with top-down cognitive control and selection of appropriate responses according to current instructions or task demands [41, 55]. In particular, a key control function of DLPFC is to resolve conflicts by allowing responses with weaker activation levels to gain priority over stronger ones under appropriate circumstances. In incompatible trials, DLPFC may therefore provide conflict resolution between action alternatives (i.e. left or right key press), through reducing activations for incompatibly-primed responses. Since AG activation negatively correlates with the subjective sense of control in incompatible trials

only, strong executive contribution of DLPFC to resolve conflict in these trials would produce a weaker activation of AG, corresponding to a greater sense of control. Overall, this suggests that AG may monitor signals of conflict resolution generated during action selection within DLPFC, to prospectively inform subjective judgements of control over action outcomes.

14.7 So is Agency Metacognitive?

Metacognition is a relatively broad term that encompasses a variety of different processes. These processes all have in common that they are second-order representations of first-order mental states. The first-order mental states that are meta-represented can range from simple forms of visual perception, in cases of perceptual confidence in detection judgements, to knowledge [42], to (perhaps) agency. The first-order mental processes and states of visual perception are clearly very different from those of action control, computationally, neutrally and phenomenally. Could there then be a single second-order process that monitors them, or does each type of first-level process require its own content-specific metacognitive monitoring circuit? A common, metacognitive monitoring circuit for all first-order processes would imply a strongly hierarchical, quasi-homuncular, cognitive organization. In contrast, independent metacognitive systems would imply a highly distributed mechanism.

To answer the question of domain-generality versus domain-specificity, we examine each alleged metacognitive domain in turn, and draw comparisons between them.

In the case of agency, Miele et al. [54] argue that JoA are metacognitive and meta-representational for two reasons. First, judgements of agency are conscious, in contrast to action monitoring and action correction, which may be unconscious [8]. Second, according to Miele et al., JoA are meta-representational. By this, it is meant that judgements of agency take first-order action representations as their content. However, this latter point seems problematic. If judgements of agency are judgements about "my actions", then Miele et al.'s point stands. However, this alternative, retrospective, inferential view would not require judgements of agency to be metacognitive. If judgements of agency were simply narrative explanations of somatosensory input ("why my body moved"), then the content is not a first-level representation of action, but, ultimately, a basic-level somatosensory signal. The critical distinction seems to be whether internal, efferent signals tag body movements as being specifically "my action". If they do, then agency is metacognitive. But, if they do not, then agency may not be based on any first-order mental states, and should therefore not be considered as a metacognitive process.

It has long been argued by ideomotor theorists that retrospective SoA is only possible because (1) the computations underlying motor control are largely unconscious, for reasons of cognitive economy, and (2) the consciously available information regarding action is largely a representation of action *effects* [38].

However, we have shown that even "vague" signals such as selection fluency, which do not produce conscious phenomenology in the conventional sense, can nevertheless participate in distinctive, first-person experiences, such as SoA. The results described above show that one can experience the cause of a conflict arising during action planning/selection even though the cause behind this conflict cannot be fully represented (or defined, or named, or even accurately identified). In other words, such conflict may elicit the feeling "that something is wrong", without (necessarily) leading to awareness or knowledge about *what* is wrong (see page 331 of this chapter). Similarly, fluent action selection appears to contribute to the "buzz" of agency, because the agent *just knows* what to do. The vagueness of these feelings is interesting: because one cannot perfectly, and consciously, represent what causes this experience of fluency or conflict, the feeling is easily mistaken for something else. In our case, selection fluency is interpreted or experienced as real agency, and selection conflict is experienced as reduced agency.

This aspect of agency experience is clearly metacognitive in the sense that it is driven by the first-level motor signals associated with planning and selecting an action. However, the prospective SoA that we have described refers to these signals in a purely experiential, rather than in a representational, format. Put differently, SoA is second-order (in the sense that it directed to identifiable first-level signals), but it is not a second-order *representation*, because the specific first-level content does not form part of the second-level phenomenology. In the terms of Muñoz [58], action selection fluency contributions to prospective agency would fit better with a "control" theory of metacognition than with a "meta-representational" theory.

14.7.1 Metacognition of Agency Versus Perceptual Metacognition

Because metacognition can be so clearly defined in the domain of perception, it is useful to compare perceptual metacognition and agency metacognition. Muñoz [58] argued that a second-order (or metacognitive) representation requires "the self attribution of a mental concept together with a first-order representation". This is easily operationalised, in the case of perceptual metacognition, by the example "I know that I see that it rains". In this example, seeing that it rains is a first-order representation of an event in the external world. "I know that I see" is, in turn, the second-order representation. However, this definition of metacognition will not do for agency, because of some specific peculiarities of action signals.

First, agency judgements do not directly represent action fluency signals, at least not in the same way that "I know that I see" represents visual signals. In particular, action fluency signals lack strong phenomenology [47], and do not form a clear first-order representation. People commonly act without being fully aware

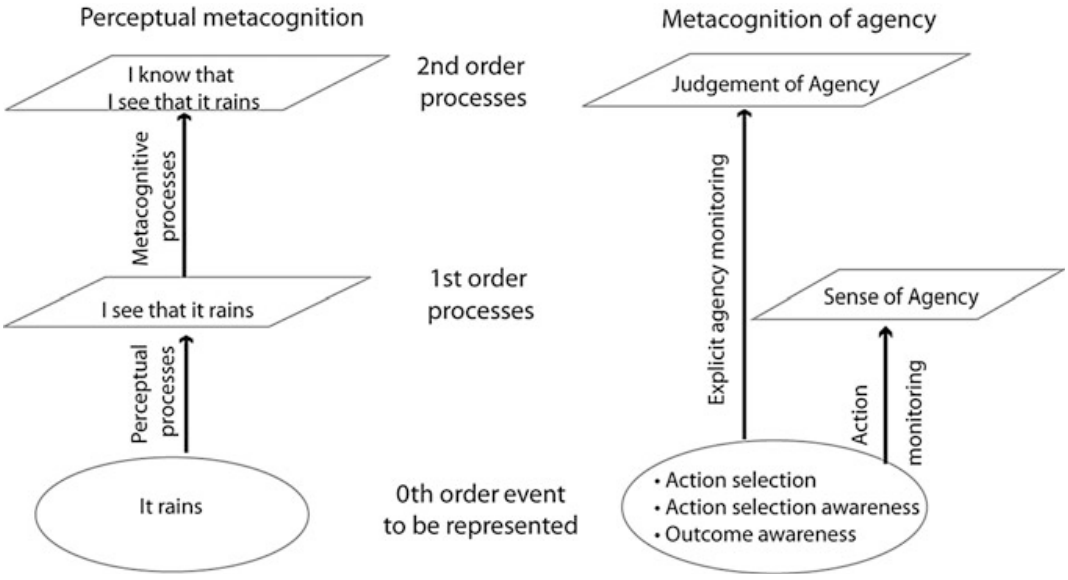

Fig. 14.4 Comparison between perceptual metacognition (*left panel*) and metacognition of agency (*right panel*). In metacognition of perception, the first-order processes are thought to be accessible to hierarchical second-order metarepresentational processes. The key feature of agency, on the other hand, is the parallel existence of two monitoring processes: A first-order process of action monitoring, and a second-order process of explicit agency monitoring both depend on the same zero-th order events, but via dissociable paths. The two paths may contribute to SoA and to JoA, respectively (see text for full details)

of their actions, leading to the textbook observation that action control is often automatic [12]. In contrast, pace the special case of blindsight, vision often provides strong phenomenology with clear first-order representational content that can be monitored, evaluated and described. To summarise, we suggest that basic-level sensorimotor signals may be processed in either or both of two dissociable ways (see schematic Fig. 14.4). They may be used by first-level processing for action monitoring, or by second-level processing for agency judgement. However, these two routes are independent, and dissociable. Unconscious adjustment of actions demonstrates the possibility of first-level processing without metacognitive SoA [8, 40]. Anosognosia for hemiplegia provides an unusual case of SoA for actions where first-level processing is effectively absent because of primary sensorimotor damage.

There is therefore a dissociable parallelism between SoA (the 'first-order' signal) and JoA (the 'second-order' signal). This implies that the two cognitive processes, namely unconscious movement monitoring and conscious agency evaluation depend on the same underlying signal, but not on each other. This is strikingly opposite to what happens in cases of perceptual metacognition, in which the hierarchically organised second-order process is formed by directly accessing the first-order process.

14.8 Implications

The nature of intentional action and agency are hotly disputed. This may be because the societal importance of these concepts is so widely recognised. Thus, individuals are responsible for their own actions, and outcomes of these actions, before society and before the law. Further, the legal concept of *mens rea* implies that individuals consciously intend particular outcomes, and that their actions realise those intentions. That is, legal responsibility depends on an SoA, which is present as part of the intentional generation of action, and at the time of controlling one's actions. However, this view has recently come under attack from two quite distinct forms of determinism. First, neurobiological determinism holds that conscious intentions are not directly controlled by persons, but are rather conscious consequences of unconscious neural events in the brain that prepare actions. Holding people responsible for unconscious, neurobiological events seems at odds with traditional ideas of 'free will' on which legal responsibility is based [36]. A second, rather different version of determinism is equally problematic for traditional ideas of legal responsibility. Social-psychological determinism suggests that people's 'voluntary' behaviour is in fact caused by subtle, often social influences of which they may be quite unaware [2].

The questions of free will, determinism and agency have been debated many times. Here, we have identified a prospective aspect of SoA, based on experimental analyses. Therefore, we simply ask, what implications does a prospective SoA have for the ideas of voluntary action and legal responsibility. First, our work shows that frontal executive processes for planning action are involved in the SoA. Our work therefore supports the idea that people are aware of actions and action outcomes (just) before they act. The neurobiological machinery that underlies planning and volition can also process action outcomes. On the other hand, our work clearly shows that this system can be driven by subliminal primes. In that sense, it is not strictly voluntary, in the sense that intentional action selection is not truly endogenous, but driven by an *external* prime.

14.9 Conclusions

To summarise, experimental analyses of responsibility suggest that the sense of being in control of one's own actions, and through them the external world, can be studied experimentally. We have distinguished between attributional and instrumental aspects of SoA. Most research to date has focused on neural mechanisms that match the predicted and actual consequences of action, and these mechanisms can be used for computing either instrumentality or attribution. These mechanisms are necessarily reconstructive, since they rely on delayed action consequences. We have argued that SoA also depends on a prospective aspect, in which fluent selection between alternative actions in the frontal cortex is monitored by parietal

mechanisms at the time of action selection itself. A component of agency is therefore computed in advance of action execution, based on a purely internal signal. This aspect of agency must, in our view, be metacognitive, but it is an experiential rather than a judgemental form of agency. Finally, we have suggested a peculiarity of agency judgement, lacking in perceptual judgement. Specifically, the 'automatic' nature of action processing means that first-level processing of action signals can occur without second-level, metacognitive, explicit self-attributive JoA. Thus, voluntary actions are generally accompanied by prospective SoA, which may be termed metacognitive. People may also make retrospective judgements of agency, which may or may not be metacognitive, depending on whether they are simply inferences about action events, or depend on internal action signals. The implications of prospective agency for voluntary control of action and legal responsibility require future research.

Acknowledgments PH was supported by an ESRC Professorial Fellowship, by EU FP7 project VERE (WP8), and by ERC Advanced Grant HUMVOL.

VC was supported by a postdoctoral bursary from the Fyssen foundation, and by EU FP7 project VERE (WP8).

References

1. Aarts H, Custers R, Wegner DM (2005) On the inference of personal authorship: enhancing experienced agency by priming effect information. Conscious Cogn 14:439–458
2. Ackerman JM, Nocera CC, Bargh JA (2010) Incidental haptic sensations influence social judgments and decisions. Science 328:1712–1715
3. Banks WP, Isham EA (2009) We infer rather than perceive the moment we decided to act. Psychol Sci 20:17–21
4. Berti A, Bottini G, Gandola M et al (2005) Shared cortical anatomy for motor awareness and motor control. Science 309:488–491
5. Blakemore S-J, Frith CD, Wolpert DM (2001) The cerebellum is involved in predicting the sensory consequences of action. NeuroReport 12:1879–1884
6. Blakemore S-J, Sirigu A (2003) Action prediction in the cerebellum and in the parietal lobe. Exp Brain Res 153:239–245
7. Bobak M, Pikhart H, Rose R et al (2000) Socioeconomic factors, material inequalities, and perceived control in self-rated health: cross-sectional data from seven post-communist countries. Soc Sci Med 51:1343–1350
8. Castiello U, Paulignan Y, Jeannerod M (1991) Temporal dissociation of motor responses and subjective awareness a study in normal subjects. Brain 114:2639–2655
9. Chambon V, Haggard P (2013) 14 Premotor or Ideomotor: how does the experience of action come about? Action Sci Found Emerg Discipl 359
10. Chambon V, Haggard P (2012) Sense of control depends on fluency of action selection, not motor performance. Cognition
11. Chambon V, Wenke D, Fleming SM et al (2013) An online neural substrate for sense of agency. Cereb Cortex 23:1031–1037
12. Charles L, van Opstal F, Marti S, Dehaene S (2013) Distinct brain mechanisms for conscious versus subliminal error detection. NeuroImage
13. David N, Cohen MX, Newen A et al (2007) The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. Neuroimage 36:1004–1014

14. David N, Newen A, Vogeley K (2008) The "sense of agency" and its underlying cognitive and neural mechanisms. Conscious Cogn 17:523–534
15. David N, Stenzel A, Schneider TR, Engel AK (2011) The feeling of agency: empirical indicators for a pre-reflective level of action awareness. Front, Psychol 2
16. De Biran M (1841) Oeuvres philosophiques. Ladrange
17. Dehaene S, Naccache L, Le Clec'H G et al (1998) Imaging unconscious semantic priming. Nature 395:597–600
18. Desmurget M, Reilly KT, Richard N et al (2009) Movement intention after parietal cortex stimulation in humans. Science 324:811
19. Desmurget M, Sirigu A (2009) A parietal-premotor network for movement intention and motor awareness. Trends Cogn Sci 13:411–419. doi:10.1016/j.tics.2009.08.001
20. Engbert K, Wohlschläger A, Haggard P (2008) Who is causing what? The sense of agency is relational and efferent-triggered. Cognition 107:693–704
21. Farrer C, Franck N, Georgieff N et al (2003) Modulating the experience of agency: a positron emission tomography study. Neuroimage 18:324–333
22. Farrer C, Frey SH, Van Horn JD et al (2008) The angular gyrus computes action awareness representations. Cereb Cortex 18:254–261
23. Farrer C, Frith CD (2002) Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. Neuroimage 15:596–603
24. Fink GR, Marshall JC, Halligan PW et al (1999) The neural consequences of conflict between intention and the senses. Brain 122:497–512
25. Fogassi L, Ferrari PF, Gesierich B et al (2005) Parietal lobe: from action organization to intention understanding. Science 308:662–667
26. Fotopoulou A, Tsakiris M, Haggard P et al (2008) The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. Brain 131(12):3432–3442
27. Fourneret P, Jeannerod M (1998) Limited conscious monitoring of motor performance in normal subjects. Neuropsychologia 36:1133–1140
28. Franck N, Farrer C, Georgieff N et al (2001) Defective recognition of one's own actions in patients with schizophrenia. Am J Psychiatry 158:454–459
29. Fried I, Mukamel R, Kreiman G (2011) Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. Neuron 69:548–562
30. Frith CD, Blakemore SJ, Wolpert DM (2000) Abnormalities in the awareness and control of action. Philos Trans R Soc Lond B Biol Sci 355:1771–1788
31. Gallagher S (2004) Neurocognitive models of schizophrenia: a neurophenomenological critique. Psychopathology 37:8–19
32. Galvin SJ, Podd JV, Drga V, Whitmore J (2003) Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychon Bull Rev 10:843–876
33. Garbarini F, Rabuffetti M, Piedimonte A et al (2012) "Moving" a paralysed hand: bimanual coupling effect in patients with anosognosia for hemiplegia. Brain 135:1486–1497. doi:10.1093/brain/aws015
34. Gergely G, Bekkering H, Király I (2002) Developmental psychology: rational imitation in preverbal infants. Nature 415:755
35. Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. Nat Neurosci 5:382–385
36. Haggard P, Libet B (2001) Conscious intention and brain activity. J Conscious Stud 8:47–64
37. Haggard P, Chambon V (2012) Sense of agency. Curr Biol 22:390–392
38. Hommel B, Musseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–877
39. Hume D (1978) A treatise of human nature [1739]. Br Moralists 1650–1800
40. Johnson H, van Beers RJ, Haggard P (2002) Action and awareness in pointing tasks. Exp Brain Res 146:451–459
41. Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302:1181–1185. doi:10.1126/science.1088545