Sense of agency in the human brain

Patrick Haggard

Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London, WC1N 3AR, UK

Correspondence to: p.haggard@ucl.ac.uk

Abstract| In adult life, people normally know what they are doing. This experience of controlling one's own actions and, through them, the course of events in the outside world is called the 'sense of agency'. It forms a central feature of human experience; however, the brain mechanisms that produce the sense of agency have only recently begun to be investigated systematically. This recent progress has been driven by the development of better measures of the experience of agency, improved design of cognitive and behavioral experiments and a growing understanding of the brain circuits that generate this distinctive but elusive experience. The sense of agency is a mental and neural state of cardinal importance for human civilization because it is frequently altered in psychopathology and because it underpins the concept of responsibility in human societies.

The sense of agency is the feeling of making something happen. It is the experience of controlling one's own motor acts and, through them, the course of external events. The term 'sense of agency' has also been used with a rather different meaning in both social science[1] and computational[2] literatures, where it typically designates a felt capacity to act (self-efficacy), without reference to any specific motor act. Here, however, I focus on the experience that occurs before, during and after actual muscular movement, rather than on beliefs or facts about potential actions. Thus, I use the term sense of agency to refer to an experience that accompanies the performance of a specific motor act.

Many voluntary actions are 'phenomenally thin'[3], because the experiences that accompany them are not particularly vivid. However, the conjunction of several different aspects of this experience is normally jointly sufficient to produce a feeling of control over what one is doing, even if one is doing it 'automatically'[4]. The neural computations that produce this experience are so efficient and so familiar that our sense of agency can seem minimal and banal. However, a simple example demonstrates the importance and careful construction of the sense of agency. When it gets dark, I may reach out to switch on the light, perhaps barely aware that I am acting at all. However, if my hand fails to touch the switch, or if the light fails to come on (or if someone else switches the lights on just before I do using another switch) I will experience a striking conflict and violation of expectations as a result of the mismatch between the intended and actual result of the action. In this scenario, the normal experience of fluently controlling the environment is suddenly interrupted as the sense of agency is lost.

Sense of agency underpins many important features of human societies. In law, for example, criminal responsibility requires not only that an agent perform a specific motor action, but also that they 'know the nature and quality of the act". This implies that the agent should experience a sense of agency with respect to their action[5,6]. Many technologies, from simple tools to social media interfaces, can extend the sense of agency from the experience of controlling the immediate environment through one's own movements to controlling much wider environments, or even virtual environments. In a major epidemiological study of work and well-being, a strong sense of agency (gained, for example, by making one's own decisions rather than executing routine tasks) was identified as a major determinant of health[7]. Disruptions of the sense of agency (in movement disorders or as a result of psychopathology, for example) have major implications for quality of life[8].

Perhaps because of its pervasive and foundational quality, sense of agency has received surprisingly little attention in cognitive neuroscience research until recent times. However, in the last two decades, neuroscientists have made greater efforts to understand the brain processes that produce sense of agency. In this Review I describe how sense of agency can be operationally defined and measured in experimental laboratory settings. I consider the different signals and cognitive processes that generate a sense of agency over one's own actions and outcomes, along with the specific brain areas and circuits that implement these computations. Disruption of these circuits may explain the altered sense of agency that frequently accompanies psychopathology. I end by discussing the importance of sense of agency for normative concepts at the level of entire societies, such as legal responsibility.

Most previous psychological and neuroscientific studies have used explicit judgements attributing actions to particular agents to study sense of agency[9,10]. However, in this review I focus primarily on the experience of instrumental control over an external object or event, rather than the attribution of authorship, and on implicit measures of sense of agency.

Defining the sense of agency

Most, but not all, actions are accompanied by characteristic subjective experiences that vary in their content and their salience. These may include the experiences of intending to act, of choosing to make one particular action rather than another and of initiating or triggering the action. These types of experience are essentially cognitive and have been linked to action preparation in the frontal lobes and to a motor command from the primary motor cortex[11,12]. They can be classed as 'central' experiences. In addition, the sense of agency typically also involves a further class of experience that is associated with the body actually moving and is relayed by the activity of peripheral somatosensory receptors. Interestingly, involuntary movements (such as reflexes and movements evoked by brain stimulation) typically produce these 'peripheral' experiences but not central experiences. Such involuntary movements are never accompanied by a sense of agency, although they are generally accompanied by what philosophers have called a sense of 'ownership'[13]. Ownership refers to the feeling of 'mineness' – that is, the feeling that an object (for example a body part or a mental state) is specifically linked to one's self [13].

Sense of agency thus seems to involve both ownership with respect to an experience of body movement ('my body moved'), and the cognitive experience of voluntary control over that body movement ('I voluntarily made it move'). The latter element has been described as the experience of oneself being the 'source of the action'[14]. This definition suggests a sense of ownership of the voluntary motor command itself, recalling both classical notions of will and neurophysiological descriptions of internally-generated action[15]. As I will describe below, recent neuroscientific evidence confirms the key role of brain circuits for voluntary action in producing a sense of agency.

The core of sense of agency, therefore, is the association between a voluntary action and an outcome. Interestingly, the importance of volition for agency is also recognized by the 'voluntary act condition' in criminal law. According to this, individuals can only be responsible for their own voluntary actions and not for their reflexes, sneezes or similar movements[16] (BOX 1). Wittgenstein famously asked "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"[17]. One possible answer might be: a sense of agency. Psychologists have noted that the sense of agency is strongest when there is a strong motivation to act, a clear action goal and a specific cortical motor command that initiates the action[18]. However, the precise definition of a voluntary action remains controversial. Some neuroscientists suggest that a voluntary action is one that requires neural activity in cognitive-motor areas[19]. Others have eschewed the concept of voluntariness altogether, and have instead defined a class of endogenous, or 'internally-generated', actions that contrast with reactions to an external stimulus[11]. Philosophers have used quite different criteria: for example, some insist that voluntary actions can only be defined based on the subjective knowledge of the agent[20]. A full discussion of the neuroscience of volition can be found elsewhere[11,21,22].

A final and important aspect of sense of agency is the experience of how one's action impacts the external environment. Through their actions, humans and other animals can transform the world around them, and also experience how they have transformed it.

Measuring the sense of agency

Any scientific account of sense of agency requires some way to measure it. The simplest measure of an individual's sense of agency is the answer to the explicit question "Did you do that?". Making such a judgement requires one to attribute sensory events to one's own intentional action, rather than some other cause (FIG. 1). This process thus resembles mirror self-recognition[23,24]. Many experimental studies that have used explicit measurements of sense of agency are fundamentally social, in the sense that agency is attributed either to oneself or to another person. As noted by Marc Jeannerod, in this type of task the brain networks for sense of agency function as a 'Who' system[25].

Studies using explicit agency judgements show a consistent cognitive bias: there is a tendency to overestimate one's own agency and to misattribute to oneself events that are unrelated to one's own action[26,27]. Strikingly, this bias is stronger when the outcome of an action is positive, rather than neutral or negative, suggesting a powerful 'self-serving' mechanism through which positive affect influences sense of agency[1]. Explicit judgements of agency in social situations may be particularly distorted by such secondary gain. For example, political leaders consistently attribute economic upturns to their own political actions[28].

Given these limitations, it has been proposed that sense of agency should instead be investigated by implicit measures. Although explicit judgements of agency are rare in everyday life, we experience a clear feeling (or 'buzz') of agency during everyday actions, even when no evaluation or judgement is required[29]. Implicit measures aim to capture this feeling without requiring people to explicitly think about agency or control. They thus potentially avoid some of the cognitive biases, and desirability effects that affect explicit judgements. Interestingly, implicit and explicit measures tend to be only weakly correlated across individuals[30], although both appear to be sensitive to factors such as reward (including self-serving bias as described above)[31,32], action–outcome interval[33], and the actual degree of instrumental control[33].

One putative implicit marker of agency focusses on distortions of time perception. Programming actions[34], executing actions[35] and predicting their outcomes[36] all influence time perception. In experiments focusing on the ‘intentional binding’ effect, participants are asked to report the perceived time of either a voluntary action or of a subsequent sensory event (such as a tone). It has been shown that voluntary actions, but not involuntary movements, are perceived as shifted in time towards their subsequent outcomes and that the outcomes themselves are perceived as shifted towards the voluntary actions that caused them, in comparison to control conditions in which the action and outcome occur independently (FIG. 2a)[37]. As a result, sense of agency can be quantified as a compression of the perceived interval between action and outcome (FIG. 2b)[38]. Many factors influence time perception, including attention, causality, pharmacological agents and adaptation. Therefore, shifts in time perception are not diagnostic of a sense of agency. However, a difference in intentional binding between two appropriately chosen conditions can potentially be interpreted as a difference in sense of agency.

The experience of instrumental control that is typically investigated in experiments such as intentional binding is largely independent of the process of social attribution to agents and explicit judgement. To take an example from daily life, someone cycling up a hill may experience a strong sense of agency that is based on perceiving how their actions affect the speed of the bike. This process does not invoke any other agent, any social aspect or any explicit propositional judgement.

The two different approaches to the measurement of agency allow us to investigate an important distinction in the psychology of action between one’s own instrumental agency (self to world control) and social attribution (self or other). This distinction has long been recognized in philosophy and psychology[39], but it remains unclear whether social attribution of actions to agents depends on an antecedent ability to compute and perceive instrumental agency, or *vice versa*. This review focuses on the sense of agency associated with instrumental control, as distinct from the notion of agency based on social attribution.

Cognitive processes that drive agency

Volition and action preparation

As noted above, a genuine sense of agency clearly requires some internal state of volition, conation, or 'urge'. Involuntary movements, such as those caused by brain stimulation or passive displacement of a body part, do not produce a sense of agency. For example, studies of the intentional binding effect show that involuntary movements produce less binding than voluntary actions, or even reverse the effect entirely[31, 40].

Recognizing the distinctive neural events that accompany volitional, as opposed to involuntary movements is therefore important for our understanding of sense of agency and may also be an important element in learning voluntary control of movement[41] (BOX 2). Preparatory activity in cognitive-motor areas has been considered characteristic of voluntary action. For example, the readiness potential, a characteristic slow negative EEG potential that occurs prior to movement, has classically been taken as a marker of volition[42] (but for a contrary view see[43]). In one recent study, participants who showed a strong early rise in the readiness potential also showed stronger intentional binding between actions and outcomes, implying an increased sense of agency[44]. According to this view, the cognitive preparation that precedes voluntary action may also contribute to sense of agency over an outcome. However, mere temporal coincidence between an involuntary movement evoked by transcranial magnetic stimulation (TMS), and ongoing preparation of voluntary action was not sufficient to produce intentional binding[45]. The cognitive preparation of an action plan needs to match the muscular movement precisely in order to elicit an experience of agency.

Action selection: choosing what to do

Choosing between alternative possible actions strongly influences sense of agency[46]. Importantly, this influence must be prospective, because the processes that select between alternative actions necessarily precede movement and outcome. During action selection, the dorsolateral prefrontal cortex (DLPFC) is thought to assemble a 'response space', a set of alternative possibilities from which the desired action must be chosen[47]. Some studies suggest that the DLPFC not only represents possible alternative actions, but also houses the process of selecting between them[48]. However, medial frontal areas have also been argued to contribute to the action selection process[49,50].

Several lines of evidence suggest that the processes of selecting between alternative actions in the frontal lobe strongly contribute to sense of agency. A recent paper[51] meta-analyzed seven studies that combined anodal transcranial direct current stimulation (TDCS) over the left DLPFC with intentional binding measures. TDCS enhanced intentional binding only in those studies in which participants themselves selected between alternative actions.

Action selection processes could affect sense of agency in several ways. When alternative actions consistently cause specific corresponding outcomes then the ability to select which action to make gives the agent some actual statistical control over the outcome, thus increasing the sense of agency by increasing the likelihood that the outcome will match that desired. For example, in one study a specific electroencephalographic (EEG) component (the auditory N1) that was evoked by a tone was more strongly attenuated when the tone reliably corresponded to whichever of four actions participants had selected, compared to when it did not[52]. Such sensory attenuation is thought to occur when a prediction of the outcome of one's action makes reafferent perception of the actual outcome redundant. The exact relation between sensory attenuation and sense of instrumental agency remains unclear (see below), but this finding demonstrates the important relation between action selection and predicting action outcomes.

Action selection processing might also directly influence sense of agency, irrespective of the relation between the selected action and its outcome. According to this view, fluent action selection (which corresponds to certainty about which action to make) boosts the feeling of control over the outcome of the action that is made. Several recent studies used subliminal visual priming to demonstrate that sense of agency depends on metacognitive signals generated during action selection. For example, briefly presenting and then masking a right-pointing arrow can reduce manual reaction time to a subsequent, supraliminal right-pointing 'target' arrow and increase reaction time to a left-pointing target arrow. Unseen primes can also bias 'free' choices between left and right hand responses[53]. Interestingly, the same primes can influence explicit judgements about the sense of agency[54]. Congruent priming boosted explicit judgements of agency over a visual event that occurred as an outcome of the manual response, relative to incongruent priming. Crucially, the primes did not directly predict the identity of the visual outcome, and further experiments ruled out the possibility that participants simply monitored their reaction times, and used motor performance as a proxy for agency[55]. Thus, action selection processes appeared to generate an internal, metacognitive signal indicating the level of fluency or conflict involved in selecting an action from the response space. These action selection signals contributed to judgements of agency, even though the difficulty or otherwise of action selection was entirely independent of the actual statistical relation between action and outcome. The contribution of action selection fluency to sense of agency is, in a sense, illusory. Just knowing what action to make does not guarantee the action outcome. However, as with many illusions, it may be a rule of thumb that normally works for typical agency situations.

The comparator model of agency

As described above, a sense of agency is generated when voluntary actions match outcomes. In computational models of motor control (FIG. 3), motor commands are used to predict the sensory consequences of action[56]. This prediction is thought to involve passing an efference copy of the motor command to a ‘forward model’ (also known as an ‘internal predictive model’) of the moving body part[57]. Sensory information about the body and the environment is then compared with the sensory feedback that would be predicted given the motor command. The result of this comparison is known as a prediction error. For example, when the brain sends the motor command to reach for the light switch, one might predict the resulting movement of the arm and also that the lights will come on. If the arm does not move in the appropriate way, the motor control system must update or alter the motor command to achieve the goal of switching the lights on.

Comparator models were originally developed to explain how the brain monitors and corrects goal-directed movements. However, the same models have also been used to explain the sense of agency. If an event is caused by one’s own action (and if the internal predictive model is correct), the actual feedback corresponds exactly to the prediction, and the result of the comparison is zero; otherwise the result is a non-zero prediction error.

According to this view, people have a sense of agency over events that can be predicted given their motor commands. In one series of studies, the comparator model was used to explain how people attribute visual feedback to their own action, or to the action of another agent (FIG. 1). The comparator model suggests that people compare the visual feedback predicted from their own motor commands with the movement they see[9,58]. Any mismatch in this comparison justifies the attribution "that's not me" implying a reduced or absent sense of agency[24]. Temporal cues can play a particularly important role in action attribution[59]. The time of initiating an action allows a precise prediction about the time of the outcome[60,61]; thus, any temporal mismatch in the visual feedback of actions reduces the sense of agency.

Interestingly, according to the comparator model, sense of agency is caused by the lack of any prediction error, implying absence of any signal at the output of the neural comparator. These models may therefore successfully explain the phenomena of 'non-agency', such as the striking feeling that "I didn't do that, something's going wrong!") when actions fail and generate a prediction error signal. However, they may be less convincing as explanations of the 'buzz' of agency in routine, successful action[29], since the model does not generate any neural signal that could cause this experience. Furthermore, comparator models attenuate or even entirely suppress, perception of action outcomes: according to these models, we only perceive what we cannot predict since the content of perception is given by the difference between actual and predicted feedback. This suppression can be functional: for example, it prevents perceptual overload during self-generated action[61]. However, it seems counterintuitive to attenuate the perception of the very goal events that are the targets of volition. If sense of agency depended only on comparator mechanisms, the experience of the fruits of one's own agency would be tantalizingly suppressed. Indeed, taking this interpretation to an extreme, human endeavor might disappear if agents never became aware of the success of their actions.

For these reasons it has been suggested that the comparator model may apply primarily to the immediate sensorimotor effects of movement[62,56] and may attenuate only high-intensity action outcomes[63]. In contrast, people can experience a sense of agency over arbitrary outcomes long after the action occurred. This suggests that cancellation of reafference against predictions cannot be the only process contributing to sense of agency.

Prospective versus retrospective agency

The different signals within the sensorimotor system that are thought to underpin sense of agency become available at different times: premotor signals (such as those associated with action selection by the inverse model) occur before action, whereas sensory feedback signals occur after the action and its outcome. Feedback signals are further affected by any delay in the causal chain between action and outcome, and by delays in receiving and processing reafferent sensory information. Comparator models deal with this temporal misalignment by delaying the output of the forward model until it can be compared to delayed outcome feedback[64]. Therefore, sense of agency can be computed only after delayed outcome feedback has reached the comparator. This implies that people experience a sense of agency only retrospectively, after the event[65]. This view receives support from the fact that explicit judgements of agency attribution are readily biased. For example, priming participants with the effect of a subsequent action increases their sense of agency over the action[66] even when the primes are below the threshold for conscious perception[67]. Strong advocates of this view claim that actions are triggered by environmental influences and premotor processes that operate largely outside of consciousness. Thus, the brain generates a sense of agency through retrospective inferences about one’s own action authorship after the fact, but does not have access to any direct readout or signal about the true origins of actions, or does not use these signals when computing sense of agency[65].

Some recent studies, however, suggest that sense of agency also depends on prospective signals. Most experimental measures of sense of agency are unable to separate the effects of prospective and retrospective components of agency. However, studies of intentional binding have distinguished between these two components by varying the probability that a voluntary action produces a tone. For example, one study showed that, when the probability that a tone would follow action is 50%, the binding of the action towards the tone is stronger in trials on which tones are actually presented, compared to trials without tones[68]. This suggests that the tone retrospectively altered the perceived timing of the action that caused it. Similar effects in other studies have been described as ‘postdiction’[69]. Interestingly, the action binding on those trials on which no tone in fact occurred depended on the probability of tone occurrence: binding was stronger on blocks where the probability of a tone following an action was high (75%) rather than low (50%)[68]. This finding suggests that an additional, prospective component of the sense of agency is present even if the outcome does not in fact occur. This experiment provided strong evidence that the experience of one’s own actions is fundamentally linked to the prediction of those actions’ outcomes, and not just to retrospectively inferring agency once outcomes are known.

Brain mechanisms underlying agency

Neuroimaging studies have played an important role in identifying the brain areas implementing the various cognitive and computational processes underlying sense of agency. These studies have consistently highlighted the role of the parietal cortex in sense of agency. Early studies that used explicit agency attribution judgements identified strong activation of the angular gyrus in the inferior parietal lobe in situations in which visual feedback was judged as unrelated to one’s own action[58,70]. A recent meta-analysis of 15 neuroimaging studies confirmed the temporoparietal junction area, including the angular gyrus, as the neural correlate of ‘non-agency’[10]. Some of the studies included were based on experimental factors known to influence sense of agency, such as delayed sensory feedback, rather than on actual subjective reports about agency. The studies also generally focused on retrospective components of the sense of agency (that is, on judging whether a stimulus was self-generated or not). Interestingly, the temporoparietal junction also responds to unexpected external sensory events in the absence of voluntary action[71]. Therefore, its activation in situations of non-agency may not reflect the process of attributing agency, but rather one possible result of that process (namely, the judgement that an event is externally-caused). Interestingly, medial and lateral prefrontal areas were also associated with non-agency over outcomes in some of the studies[10]. The only area that was consistently associated with positive self-agency was the anterior insula, possibly reflecting a general role of this area in ongoing self-awareness[72].

A recent event-related fMRI experiment, suggested that the angular gyrus may also contribute to the prospective sense of agency. In this study[73], researchers used subliminal priming to alter the fluency of action selection. Participants made left and right hand keypresses in response to arrow stimuli. Their actions lead, after a delay, to one of a number of colors appearing on the screen. An unseen, subliminal prime arrow presented just before the imperative stimulus could facilitate or hinder the participants in selecting and making the appropriate action, depending on whether the subliminal prime arrow was compatible or incompatible with the imperative stimulus. Compatible primes led to faster reactions than incompatible primes[74]. More interestingly, compatible subliminal primes also led to a stronger sense of control over the color patch caused by the participant's action. FMRI analysis focused on the brain activations associated with action selection. The angular gyrus was more strongly activated when participants gave low agency ratings, rather than high agency ratings, but only when the subliminal prime was incompatible with the supraliminal imperative stimulus. Thus, the angular gyrus may monitor signals generated by conflict arising during action selection processes in frontal areas. This hypothesis was supported by an increased functional connectivity between the angular gyrus and lateral prefrontal cortex during incompatible trials. Furthermore, TMS of the angular gyrus at the time of presentation of the primes abolished the tendency to give lower agency ratings on incompatible trials[75]. Thus, the parietal cortex may not simply match outcomes to actions retrospectively. It may also receive prospective signals about ongoing processing for the selection and initiation of action. Because this prospective component can be measured independently of actual contingencies between actions and outcomes, it could be considered metacognitive.

The contribution of frontal and prefrontal lobes to sense of agency remains less clear. These areas play a crucial role in planning and initiation of voluntary action[7]. Many experimental designs use manipulations that influence both the frontal processes that generate motor actions, and those that generate the subjective experience of agency. These designs cannot distinguish between direct effects on sense of agency, and indirect effects mediated by changes in action control. However, some recent studies combining non-invasive brain stimulation with implicit measures of sense of agency have focused specifically on the experience of agency. It was shown[76] that TDCS over the pre-supplementary motor area (pre-SMA), a medial frontal area strongly implicated in both volition and sense of agency, reduced the intentional binding effect, and particularly the binding of actions towards outcomes. In another study, theta-burst TMS over pre-SMA reduced the binding of outcomes towards actions[77]. Finally, as described above a recent meta-analysis[51] suggested that action selection processes in the DLPFC make a specific contribution to the sense of agency.

These recent studies thus provide an overall picture of the brain networks underlying sense of agency. Prospective signals alone can provide only estimates about instrumental control over external events: the chain of causation may go awry, or one may be unable to track the causal chain correctly[78]. Inferences based on retrospective signals alone appear to misclassify many events as self-caused when they are not. Therefore, a combination of prospective and retrospective signals seems necessary to reliably compute one's own agency. The parietal cortex plays a key role in monitoring multiple signals relevant to sense of agency, including signals of selection, volition and initiation generated in the frontal cortices[79]. According to one view[79], the parietal lobe acts as a comparator model, comparing intentional signals with sensory feedback signals. However, this account cannot explain why parietal activations strongly depend on the results of the comparison, rather than on the process of comparison itself. In particular, several studies showed stronger parietal involvement for a sense of non-agency than for positive sense of agency[9,58].

Comparator models of agency encourage the search for a single brain site where intention and feedback are matched. However, an alternative view treats agency as a subjective consequence of an association between action and outcome, rather than as a result of difference signals between predicted and actual outcome. Such associations might link the prospective premotor signals arising from action selection and initiation to the retrospective, perceptual signals arising from body movement and external outcomes. The key neural correlate of sense of agency might lie in the connectivity between frontal and prefrontal motor areas that initiate action, and parietal areas that underlie the monitoring of perceptual events, rather than in any single structure. A recent study on sense of agency in psychosis (see below) strongly supports this connectivity view.

Pathological sense of agency

Several neurologic and psychiatric conditions involve pathology of sense of agency. Clinical neurology broadly classifies movement disorders as either hyperkinetic or hypokinetic. A similar classification can be made for the subjective experience of action. Some conditions are hyperagentic (involving an excessive experience of one's own causation and control over events), whereas others are hypoagentic (involving reduced experience of causation and control). One account of depression notes that depressed individuals experience less control over outcomes than healthy individuals[80], but attributes this to a hyperagentic bias of healthy individuals. According to this view, which has remained controversial, depressed individuals have a realistic assessment of how limited their agentic capacity really is.

Most research on pathology of agency has focused on schizophrenia. The German psychiatric tradition uses the term *Ich-Störung* to refer to a range of disturbances of selfhood, the self-world boundary, and self-agency. For example, patients suffering from delusions of control may have the feeling that their thoughts and actions are not their own, but are instead transmitted or caused by external agents. According to one model[81], such symptoms arise from a failure to predict the consequences of self-generated actions. In the absence of appropriate predictions, sensory experiences caused by one's own actions and thought processes are not cancelled by any internal prediction, and are perceived as external events. Several lines of evidence have supported this view. Patients with schizophrenia have shown an impaired ability to detect when visual feedback of their movements had been modified[82] or was not linked to their own action[83]. Intriguingly, they also lacked the sensory attenuation of self-generated action consequences, a key feature of the comparator model of predictive motor control[84,85,86].

These results would suggest a reduced sense of agency in psychosis. However, a recent study found that the intentional binding effect in a small sample of patients with positive symptoms was stronger, rather than weaker, than in healthy volunteers[87]. Underestimation of time intervals appears to be a general characteristic of some patients with schizophrenia[88]: indeed, the patients showed a stronger binding between two successive tones, in the absence of action, than did the controls. However, the between-group difference in binding was greater for action-tone intervals than for tone-

tone intervals. In a larger, more recent study, it was shown that the main distortion of sense of agency in schizophrenia may lie not in the strength of the agency experience, but in the cognitive processes that produce it[89]. In an experiment designed to isolate the prospective and retrospective components of the sense of agency (see above and FIG. 4) the intentional binding of actions towards tones in patients with schizophrenia depended entirely on the actual occurrence of the tone. By contrast, in the healthy volunteers, action binding depended on the probability that the tone might occur, and was largely independent of whether the tone actually occurred or not on any individual trial. That is, the patients' sense of agency was based largely on retrospective reconstructions of the experience of action, driven by occurrence of the outcome, whereas the healthy volunteers' sense of agency was based on predictions about likely outcomes. Psychosis thus seemed to abolish the normal prospective element of agency that characterizes healthy adult life. In this sense, the results of binding and sensory attenuation experiments agree: in schizophrenia, the brain may be in a constant state of surprise, attempting to understand the events it has itself generated.

Society, agency and responsibility

Most human societies adhere to the concept of individual responsibility for action. This forms the basis for praise and blame, punishment and reward. Individual responsibility depends on the assumption that most, or all, individuals experience a sense of agency over their actions and outcomes. In fact, courtroom pleas of 'guilty' or 'not guilty' are explicit judgements of agency. Few mental states thus sustain such a strong social superstructure as the sense of agency. The 'voluntary act condition' in law[16] insists that an individual can only be criminally responsible for actions that they consciously decide to perform with a reasonable understanding of the likely outcome. The emerging field of 'Neurolaw' has highlighted the relationship between legal concepts of responsibility and individuals' cognitive capacity to experience agency[90]. For example, the law acknowledges diminished responsibility for homicidal action on the grounds of 'loss of control' in a small number of specific cases, such as self-defense, or persistent abuse (BOX 1). Does such 'loss of control' represent involuntary recruitment of a specific neurobiological system subserving 'flight or fight' action control, or merely a cultural expectation about when extreme behavior might actually become acceptable? These explanations need not be exclusive: the law can be seen as a normative constraint on basic neurobiological drives to action.

In other controversial cases, defendants have denied responsibility by claiming that they were only obeying orders. This implies that sense of agency for one's own actions may be reduced under conditions of coercion, which could in turn explain why people appear to comply rather readily with coercive instructions[91]. On the other hand, this defense could simply represent a disingenuous attempt to avoid blame by claiming a reduced experience of agency. Accordingly, an experimental design was recently developed in which an experimenter gave instructions to one participant to deliver visibly painful shocks to another participant (FIG. 5)[92]. Because the participants took turns in the roles of giving and receiving shocks, each had direct experience of the unpleasant effects of their own actions on their co-participant. Participants estimated the brief delay that elapsed between their action and an auditory tone that occurred synchronously with the delivery of the shock. The subjective duration of this interval provides an implicit measure of sense of agency, analogous to intentional binding, with intervals initiated by voluntary actions being perceived as shorter than equivalent intervals initiated by passive movements (see FIG. 2b[93,94]). Coercive instructions increased the perceived duration between action and tone/shock, relative to free choice, indicating that coercion reduced sense of agency. Strikingly, coercive instructions also resulted in lower event-related potential **(**ERP**)** responses to the tone, suggesting a suppression of neural processing of action outcomes. This implies that obeying instructions reduces the sense of agency, producing an experience closer to passive movement than to voluntary action.

Previous research had shown that increasing the number of action alternatives from which a participant could freely select leads to an increase in sense of agency[46]. Reduced sense of agency under coercion is consistent with the possibility that social coercion effectively constrains an individual's free choice[47]. The finding does not, of course, legitimate the defense of 'only obeying orders', but it does explain how acting under coercion can lead individuals to experience a psychological distance from unpleasant consequences of their actions. It also demonstrates how social situations can strongly influence an individual's sense of agency, and thus their feelings of responsibility.

Conclusion

The human sense of agency is not a transcendental feature of human nature, but the result of specific activity in brain circuits that underlie voluntary motor control. Frontal and prefrontal areas select and initiate intentional actions, and convey information to parietal areas that monitor intentions, actions and outcomes. This circuit operates both prospectively (in advance of actions) as well as retrospectively (to monitor whether an action has achieved the intended outcome). The computational flexibility of this circuit probably underlies the human capacity to develop a sense of agency when using advanced technologies that involve arbitrary and indirect relations between actions and outcomes. Recent psychological research has emphasized that sense of agency can result from post hoc inference, and is prone to illusions[65]. However, this argument has perhaps been overstated. In fact, the pre-motor signals related to selection and initiation of the motor command play an important role in sense of agency for the thousands of simple instrumental actions that adult humans execute each day. The distinctive experience of initiating a voluntary motor action appears to be necessary, though not sufficient, for sense of agency in normal circumstances.

Concepts of individual responsibility for action play a crucial role in the systems of law that underpin functional human societies. Individual responsibility relies heavily on brain mechanisms underlying sense of agency. Growing neuroscientific knowledge about the neural mechanisms of agency will lead to increasingly sophisticated manipulations of the subjective experience of agency, both technologically, and socio-politically. Therefore, this area of neuroscience has profound ethical implications, at the level of the individual, the society, and indeed our species as a whole.

References

1. Bandura, A. Self-efficacy mechanism in human agency. Am Psychol 37, 122–147 (1982).
2. Friston, K. et al. The anatomy of choice: active inference and agency. Front Hum Neurosci 7, 598 (2013).
3. Tsakiris, M. & Fotopoulou, A. in Decomposing the Will 103–117 (Oxford University Press, 2013).
4. Broadbent, D. E. Task combination and selective intake of information. Acta Psychol (Amst) 50, 253–290 (1982).
5. 114.R v M'Naghten. UKHL J16 (1843).
6. Garrison, A. H. The history of the M'Naghten insanity defense and the use of posttraumatic stress disorder as a basis of insanity. American Journal of Forensic Psychology 16, 39–88 (1998).
7. Marmot, M. G. et al. Health inequalities among British civil servants: the Whitehall II study. The Lancet 337, 1387–1393 (1991).
8. Frith, C. D., Blakemore, S. J. & Wolpert, D. M. Abnormalities in the awareness and control of action. Philos Trans R Soc Lond B Biol Sci 355, 1771–1788 (2000).
9. Farrer, C., Bouchereau, M., Jeannerod, M. & Franck, N. Effect of distorted visual feedback on the sense of agency. Behav Neurol 19, 53–57 (2008).
10. Sperduti, M., Delaveau, P., Fossati, P. & Nadel, J. Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. Brain Struct Funct 216, 151–157 (2011).
11. Passingham, R. E. & Wise, S. P. The neurobiology of the prefrontal cortex : anatomy, evolution, and the origin of insight. (Oxford Univ. Press, 2014).
12. Sherrington, C. S. The integrative action of the nervous system. (Yale University Press, 1906).
13. Gallagher, S. Philosophical conceptions of the self: implications for cognitive science. Trends in Cognitive Sciences 4, 14–21 (2000).
14. Marcel, A. J. in Agency and Self-Awareness: Issues in Philosophy and Psychology (eds. Roessler, J. & Eilan, N.) 48–93 (Oxford: Clarendon Press, 2003).
15. Passingham, R. E., Bengtsson, S. L. & Lau, H. C. Medial frontal cortex: from self-generated action to reflection on one's own performance. Trends in Cognitive Sciences 14, 16–21 (2010).
16. Hart, H. L. A. in Punishment and Responsibility 90–112 (Oxford University Press, 1960).
17. Wittgenstein, L. Philosophical Investigations. (Basil Blackwell, 1967).
18. Bandura, A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 52, 1–26 (2001).
19. Fried, I., Mukamel, R. & Kreiman, G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. Neuron 69, 548–562 (2011).
20. Ford, A. in Essays on Anscombe's Intention (eds. Ford, A., Hornsby, J. & Stoutland, F.) (Harvard UP, 2011).
21. Haggard, P. Human volition: towards a neuroscience of will. Nat Rev Neurosci 9, 934–946 (2008).
22. Pacherie, E. The phenomenology of action: A conceptual framework. Cognition 107, 179–217 (2008).
23. Gallup, G. G. J. Chimpanzees: self-recognition. Science 167, 86–87 (1970).
24. Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N. & Jeannerod, M. Perception of self-generated movement following left parietal lesion. Brain 122 (Pt 10), 1867–1874 (1999).
25. Georgieff, N. & Jeannerod, M. Beyond Consciousness of External Reality: A 'Who' System for Consciousness of Action and Self-Consciousness. Consciousness and Cognition 7, 465–477 (1998).
26. Tsakiris, M., Haggard, P., Franck, N., Mainy, N. & Sirigu, A. A specific role for efferent information in self-recognition. Cognition 96, 215–231 (2005).
27. Wegner, D. M. & Wheatley, T. Apparent mental causation. Sources of the experience of will. Am Psychol 54, 480–492 (1999).
28. Horne, A. Macmillan: The Official Biography. (Macmillan, 2008).
29. Synofzik, M., Vosgerau, G. & Newen, A. Beyond the comparator model: a multifactorial two-step account of agency. Conscious Cogn 17, 219–239 (2008).
30. Dewey, J. A. & Knoblich, G. Do Implicit and Explicit Measures of the Sense of Agency Measure the Same Thing? PLoS ONE 9, e110118 (2014).
31. Yoshie, M. & Haggard, P. Negative Emotional Outcomes Attenuate Sense of Agency over Voluntary Actions. Current Biology 23, 2028–2032 (2013).
32. Takahata, K. et al. It's Not My Fault: Postdictive Modulation of Intentional Binding by Monetary Gains and Losses. PLoS ONE 7, e53421 (2012).

33. Berberian, B., Sarrazin, J.-C., Le Blaye, P. & Haggard, P. Automation Technology and Sense of Control: A Window on Human Agency. PLoS ONE 7, e34075 (2012).
34. Hagura, N., Kanai, R., Orgs, G. & Haggard, P. Ready steady slow: action preparation slows the subjective passage of time. Proc Biol Sci 279, 4399–4406 (2012).
35. Yarrow, K., Haggard, P., Heal, R., Brown, P. & Rothwell, J. C. Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. Nature 414, 302–305 (2001).
36. Morrone, M. C., Ross, J. & Burr, D. Saccadic eye movements cause compression of time as well as space. Nat Neurosci 8, 950–954 (2005).
37. Haggard, P., Clark, S. & Kalogeras, J. Voluntary action and conscious awareness. Nat Neurosci 5, 382–385 (2002).
38. Caspar, E. A., Cleeremans, A. & Haggard, P. The relationship between human agency and embodiment. Conscious Cogn 33, 226–236 (2015).
39. de Biran, P. M. The influence of habit on the faculty of thinking. (The Williams & Wilkins company, 1929).
40. Moore, J. W., Wegner, D. M. & Haggard, P. Modulating the sense of agency with external cues. Conscious Cogn 18, 1056–1064 (2009).
41. Ganos, C. et al. Volitional action as perceptual detection: predictors of conscious intention in adolescents with tic disorders. Cortex 64, 47–54 (2015).
42. Shibasaki, H. & Hallett, M. What is the Bereitschaftspotential? Clin Neurophysiol 117, 2341–2356 (2006).
43. Schurger, A., Sitt, J. D. & Dehaene, S. An accumulator model for spontaneous neural activity prior to self-initiated movement. Proc Natl Acad Sci U S A 109, E2904–2913 (2012).
44. Jo, H.-G., Wittmann, M., Hinterberger, T. & Schmidt, S. The readiness potential reflects intentional binding. Frontiers in Human Neuroscience 8, 421 (2014).
45. Haggard, P. & Clark, S. Intentional action: conscious experience and neural prediction. Conscious Cogn 12, 695–707 (2003).
46. Barlas, Z. & Obhi, S. S. Freedom, choice, and the sense of agency. Front Hum Neurosci 7, 514 (2013).
47. Fletcher, P. C., Shallice, T. & Dolan, R. J. 'Sculpting the response space'--an account of left prefrontal activation at encoding. Neuroimage 12, 404–417 (2000).
48. Rowe, J. B., Hughes, L. & Nimmo-Smith, I. Action selection: a race model for selected and non-selected actions distinguishes the contribution of premotor and prefrontal areas. Neuroimage 51, 888–896 (2010).
49. Brass, M. & Haggard, P. The what, when, whether model of intentional action. Neuroscientist 14, 319–325 (2008).
50. Rae, C. L., Hughes, L. E., Weaver, C., Anderson, M. C. & Rowe, J. B. Selection and stopping in voluntary action: a meta-analysis and combined fMRI study. Neuroimage 86, 381–391 (2014).
51. Khalighinejad, N., Di Costa, S. & Haggard, P. Endogenous action selection processes in dorsolateral prefrontal cortex contribute to sense of agency: a meta-analysis of tDCS studies of 'intentional binding'. Brain Stimul 9, 372–379 (2016).
52. Hughes, G., Desantis, A. & Waszak, F. Attenuation of auditory N1 results from identity-specific action-effect prediction. Eur J Neurosci 37, 1152–1158 (2013).
53. Schlaghecken, F. & Eimer, M. Masked prime stimuli can bias 'free' choices between response alternatives. Psychon Bull Rev 11, 463–468 (2004).
54. Wenke, D., Fleming, S. M. & Haggard, P. Subliminal priming of actions influences sense of control over effects of action. Cognition 115, 26–38 (2010).
55. Chambon, V. & Haggard, P. Sense of control depends on fluency of action selection, not motor performance. Cognition 125, 441–451 (2012).
56. Blakemore, S. J., Wolpert, D. & Frith, C. Why can't you tickle yourself? Neuroreport 11, R11–16 (2000).
57. Blakemore, S. J., Wolpert, D. M. & Frith, C. D. Central cancellation of self-produced tickle sensation. Nat Neurosci 1, 635–640 (1998).
58. Farrer, C. & Frith, C. D. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. Neuroimage 15, 596–603 (2002).
59. Stetson, C., Cui, X., Montague, P. R. & Eagleman, D. M. Motor-sensory recalibration leads to an illusory reversal of action and sensation. Neuron 51, 651–659 (2006).
60. Walsh, E. & Haggard, P. Action, prediction, and temporal awareness. Acta Psychol (Amst) 142, 220–229 (2013).

61. Williams, S. R., Shenasa, J. & Chapman, C. E. Time course and magnitude of movement-related gating of tactile detection in humans. I. Importance of stimulus location. J Neurophysiol 79, 947–963 (1998).
62. Bays, P. M., Flanagan, J. R. & Wolpert, D. M. Attenuation of self-generated tactile sensations is predictive, not postdictive. PLoS Biol 4, e28 (2006).
63. Reznik, D., Henkin, Y., Levy, O. & Mukamel, R. Perceived loudness of self-generated sounds is differentially modified by expected sound intensity. PLoS One 10, e0127651 (2015).
64. Miall, R. C., Weir, D. J., Wolpert, D. M. & Stein, J. F. Is the cerebellum a smith predictor? J Mot Behav 25, 203–216 (1993).
65. Wegner, D. M. The illusion of conscious will. (MIT Press, 2002).
66. Aarts, H., Custers, R. & Wegner, D. M. On the inference of personal authorship: enhancing experienced agency by priming effect information. Conscious Cogn 14, 439–458 (2005).
67. Gentsch, A. & Schutz-Bosbach, S. I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. J Cogn Neurosci 23, 3817–3828 (2011).
68. Moore, J. & Haggard, P. Awareness of action: Inference and prediction. Conscious Cogn 17, 136–144 (2008).
69. Eagleman, D. M. & Sejnowski, T. J. Motion Integration and Postdiction in Visual Awareness. Science 287, 2036 (2000).
70. Farrer, C. et al. Modulating the experience of agency: a positron emission tomography study. Neuroimage 18, 324–333 (2003).
71. Kincade, J. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L. & Corbetta, M. An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. J Neurosci 25, 4593–4604 (2005).
72. Craig, A. D. B. How do you feel--now? The anterior insula and human awareness. Nat Rev Neurosci 10, 59–70 (2009).
73. Chambon, V., Wenke, D., Fleming, S. M., Prinz, W. & Haggard, P. An online neural substrate for sense of agency. Cereb Cortex 23, 1031–1037 (2013).
74. Eimer, M. & Schlaghecken, F. Response facilitation and inhibition in subliminal priming. Biol Psychol 64, 7–26 (2003).
75. Chambon, V., Moore, J. W. & Haggard, P. TMS stimulation over the inferior parietal cortex disrupts prospective sense of agency. Brain Struct Funct 220, 3627–3639 (2015).
76. Cavazzana, A., Penolazzi, B., Begliomini, C. & Bisiacchi, P. S. Neural underpinnings of the 'agent brain': new evidence from transcranial direct current stimulation. Eur J Neurosci 42, 1889–1894 (2015).
77. Moore, J. W., Ruge, D., Wenke, D., Rothwell, J. & Haggard, P. Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. Proceedings of the Royal Society B: Biological Sciences 277, 2503–2509 (2010).
78. Bratman, M. E. Intention, Plans, and Practical Reason. (Cambridge University Press, 1999).
79. Desmurget, M. & Sirigu, A. A parietal-premotor network for movement intention and motor awareness. Trends Cogn Sci 13, 411–419 (2009).
80. Alloy, L. B. & Abramson, L. Y. Judgment of contingency in depressed and nondepressed students: sadder but wiser? J Exp Psychol Gen 108, 441–485 (1979).
81. Frith, C. D., Blakemore, S. & Wolpert, D. M. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res Brain Res Rev 31, 357–363 (2000).
82. Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P. & Lindner, A. Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. Brain 133, 262–271 (2010).
83. Daprati, E. et al. Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. Cognition 65, 71–86 (1997).
84. Blakemore, S. J., Smith, J., Steel, R., Johnstone, C. E. & Frith, C. D. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. Psychol Med 30, 1131–1139 (2000).
85. Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D. & Wolpert, D. M. Evidence for sensory prediction deficits in schizophrenia. Am J Psychiatry 162, 2384–2386 (2005).
86. Shergill, S. S. et al. Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. JAMA Psychiatry 71, 28–35 (2014).
87. Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M. & Franck, N. Awareness of action in schizophrenia. Neuroreport 14, 1081–1085 (2003).

88. Waters, F. & Jablensky, A. Time discrimination deficits in schizophrenia patients with first-rank (passivity) symptoms. Psychiatry Research 167, 12–20 (2009).
89. Voss, M. et al. Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. Brain 133, 3104–3112 (2010).
90. Moore, M. S. Law and psychiatry: rethinking the relationship. (Cambridge University Press, 1984).
91. Milgram, S. Behavioral Study of obedience. The Journal of Abnormal and Social Psychology 67, 371–378 (1963).
92. Caspar, E. A., Christensen, J., Cleeremans, A. & Haggard, P. Coercion changes the sense of agency in the human brain. Current Biology (2016).
93. Engbert, K., Wohlschlager, A. & Haggard, P. Who is causing what? The sense of agency is relational and efferent-triggered. Cognition 107, 693–704 (2008).
94. Wegner, D. M. & Sparrow, B. in The Cognitive Neurosciences: 3rd Edition (ed. Gazzaniga, M.) 1201–1209 (MIT Press, 2004).
95. Lacey, N. Responsibility without consciousness. Oxford Journal of Legal Studies 35, 665–696 (2015).
96. Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. 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 (Pt 3), 623–642 (1983).
97. DPP v Camplin. AC 705 (1978).
98. Casey, B. J., Jones, R. M. & Hare, T. A. The Adolescent Brain. Ann N Y Acad Sci 1124, 111–126 (2008).
99. R v Ahluwalia. Cr App R 96, 133 (1993).
100. Abramson, L. Y., Seligman, M. E. & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. Journal of Abnormal Psychology 87, 49–74 (1978).
101. Coroners and Justice Act. s52 (2009).
102. Maier, S. F. & Watkins, L. R. Role of the medial prefrontal cortex in coping and resilience. Brain Res 1355, 52–60 (2010).
103. Kaye, K. & Fogel, A. The temporal structure of face-to-face communication between mothers and infants. Developmental Psychology 16, 454–464 (1980).
104. Ruvolo, P., Messinger, D. & Movellan, J. Infants Time Their Smiles to Make Their Moms Smile. PLoS One 10, e0136492 (2015).
105. Cipriani, D. Children’s Rights and the Minimum Age of Criminal Responsibility: A Global Perspective. (Ashgate Pub., 2009).
106. Brown, C. My Left Foot (Simon & Schuster, 1955).
107. Taylor, A. H. et al. Of babies and birds: complex tool behaviors are not sufficient for the evolution of the ability to create a novel causal intervention. Proc Biol Sci 281, (2014).
108. Iriki, A. & Taoka, M. Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. Philos Trans R Soc Lond B Biol Sci 367, 10–23 (2012).
109. Iriki, A., Tanaka, M., Obayashi, S. & Iwamura, Y. Self-images in the video monitor coded by monkey intraparietal neurons. Neurosci Res 40, 163–173 (2001).
110. Hochberg, L. R. et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442, 164–171 (2006).
111. Koralek, A. C., Jin, X., Long, J. D., Costa, R. M. & Carmena, J. M. Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. Nature 483, 331–335 (2012).

Acknowledgments

This work was supported by ERC Advanced Grant HUMVOL (agreement number 323943) and by a grant from AHRC (number 162746). The article benefitted from a fellowship at the Paris Institute for Advanced Studies (France), with the financial support of the French State managed by the Agence Nationale de la Recherche, programme “Investissements d’avenir”, (ANR-11-LABX-0027-01 Labex RFIEA+).

The author would like to thank Elisa Brann for comments and assistance.

BOX 1: Agency and voluntary control in the law courts

All legal systems have some concept of criminal responsibility. This concept asserts that healthy adults have a conscious experience and knowledge of their own intentions, actions and outcomes that underpins their voluntary control of the action. Systems of law descending from Roman Law use the term *mens rea* to describe this subjective, intentional aspect of responsibility. In principle, therefore, someone who lacks a sense of agency regarding an action cannot be held criminally responsible for that action. In practice, courts are rightly skeptical when defendants claim that they did not experience what they were doing: there is an obvious secondary gain associated with such a claim and lack of volition and agency are difficult to prove objectively. The legal concept of *mens rea* also contrasts with neuroscientific views that emphasize the automatic, unconscious precursors of actions that are experienced as voluntary[95,96,43].

However, the law acknowledges some situations in which voluntary control and agency over action are reduced. Such cases can often be understood in terms of interactions between limbic and motor systems in the brain. Past homicide cases have considered whether factors such as age (and thus the maturational state of the brain) or circumstances that induce intense emotion or involve prolonged abuse can alter these limbic-motor interactions and thus influence voluntary control of action[97,98,99]. For example, prolonged abuse can induce profound changes in cognitive and behavioral capacity, similar to the 'learned helplessness' condition that is described in animal behavior research[100]. Thus homicide in such cases could viewed as analogous to a 'fight' response to threat, which may replace the more normal 'flight' response under conditions of extreme stress.

In English law, the defense of 'provocation' was recently replaced by a 'loss of control' defence[101]. It was recognized that actions following prolonged abuse (and not only actions made in the sudden heat of the moment) may not meet normal conditions of voluntary control and responsibility. Interestingly, rats with previous experience of control over a stressor acquired resilience against effects of subsequent uncontrollable stress[102]. Past history of agency may thus be strongly neuroprotective in situations of uncontrolled stress. Conversely, sustained abuse might remove the resilience which, under normal circumstances, allows a degree of self-control even in occasional situations of high stress, threat or emotion. This could provide a neural basis for considering duration of abuse as a relevant 'qualifying trigger' for actions involving loss of control. However, this area of law remains highly controversial, and the relevance of neuroscientific evidence is, correctly, a topic of intense debate.

BOX 2: Acquiring a sense of agency

A newborn infant appears to have little or no voluntary control over its actions, and little or no sense of agency. Classical theories of psychological development suggest that voluntary control of actions (notably smiling) emerges in the first months of life and is guided by the reinforcement of parental affective communication[103,104]. This observation suggests that humans learn to become agents over their own bodies, and thus over their external environment (including their caregivers). This initial learning generally predates first autobiographical memories; however, the learning process continues well into adulthood. Indeed, some legal systems acknowledge fully-developed sense of agency and responsibility only after 18 years of life[105]. The case of Christy Brown, who was born with cerebral palsy and was only able to move his left foot, offers an interesting insight into the transformative power of sense of agency. In his autobiography, Brown recalls the moment of his first instrumental action – using his left foot to seize a piece of chalk from his sister's hand, and draw a squiggle. He describes this first experience of agency as the starting point of a virtuous cascade of escalating achievement that was recognized and encouraged by others, and culminated in a successful artistic career[106].

The same progressive trajectory of agency occurs both in individual learning, and on an evolutionary scale. For example, tool-use can be seen as an evolutionary adaptation that extended the range of an animal's agency[107,108]. Humans, and perhaps some other primates, can learn and exploit arbitrary associations between actions and outcomes. As a result, they retain a clear sense of agency, even when using sophisticated technologies to act at a distance[109]. The recent development of brain-machine interfaces offers novel examples of learning a sense of agency. Paralyzed individuals can learn to control a neuroprosthesis by thought[110]. Anecdotal evidence suggests these patients report a feeling of agency for actions using the prosthesis. Although no human studies have yet systematically explored how this experience emerges during learning, rodent models suggest the plasticity of corticostriatal connections plays a key role in acquiring neuroprosthetic agency[111]. Once the basic cognitive capacity of instrumental control develops, it appears to transfer successfully to novel brain codes for action, as well as to arbitrary outcome events. The dramatic pace of human technological achievement suggests that the cognitive mechanisms underlying instrumental agency may be almost unlimited in scope.

Figure captions

Fig 1 Explicit judgments of agency

Action-recognition experiments can be used to examine explicit judgments of agency. In a typical example of this type of experiment, participants are asked to judge whether a video that they are watching shows their own hand movements or those of another person. Participants are asked to make a specific pattern of hand movement. A screen is connected to a video switch (controlled by the experimenter), allowing the participant to see either their own hand or the hand of an experimenter wearing an identical glove. The experimenter performs either the same hand movement as the subject, or a different hand movement. If the participant reports that they are viewing their own hand action, they attribute authorship of the viewed action to themselves. Figure adapted, with permission, from [24]

Fig 2 Measuring sense of agency implicitly

a| A schematic showing how the intentional binding task can provide an implicit measurement of sense of agency[45]. In this task, participants view a small rotating clock hand. They make a self-paced button press action, which triggers a tone 250 ms later. Thus, participants are expected to experience a sense of agency over the tone. They report the time at which they pressed the button or at which they heard the tone, in separate blocks of trials. First, the participants are asked to estimate the clock time at which they made the action or heard the tone in separate baseline conditions in which only the action or only the tone occurs. Next, in the experimental conditions, the participant themselves causes the tone by their own action, and judges either the time of the action or the time of the tone. In this case, the perceived times of the action and tone shift closer to each other (indicated by the dashed arrows), relative to the baseline conditions. This produces an intentional binding effect. Replacing the intentional action with a physically similar but involuntary movement (a twitch evoked by transcranial magnetic stimulation (TMS) of the motor cortex) has the opposite effect. Thus, sense of agency can be measured implicitly and quantitatively using the magnitude of perceptual shifts in action and tone timing. b| In an interval estimation paradigm, participants give an absolute verbal estimate of the delay (in ms) between a keypress action and a tone that follows after a short but somewhat variable delay. Lower estimated delays correspond to a greater sense of agency[93]. Thus, estimates are shorter for a voluntary keypress than they are in a control condition in which the finger is passively pushed onto the response key. The actual intervals between keypress and tone are randomized, but are always less than one second to prevent counting. Part b adapted, with permission, from[38].

Fig 3 The comparator model for neural control of action and agency

In the comparator model of action control, an action begins with an intention or desired goal state. An inverse model computes the motor command required to achieve the goal state (or at least approach it) and generates the motor command that will drive the action. A forward model uses a copy of the current motor command (known as an efference copy) to predict the likely sensory consequences of the command[56]. This prediction is compared with sensory feedback signals that provide information about the ongoing action, and about its effects on the external environment. The result of the comparison can be used in three ways: to adjust the current motor command (1), to attribute agency for actions and environmental events (2; if the comparator gives a result of 0 then the event is caused by one's own action) and to attenuate predictable, self-produced sensations (3).

Fig 4 Prospective and retrospective agency.

Schematics illustrating the results of an experiment that used the intentional binding effect to distinguish between prospective and retrospective components of the sense of agency. a| Voluntary actions were followed by tones in 50% of trials (trials chosen at random). In healthy volunteers, the perceived time of the action (blue rectangles) was independent of whether the tone occurred or not. In patients with schizophrenia, the perceived time of the action showed a general shift towards the time of the tone (grey rectangles), perhaps indicating poor attention to time. Crucially, the perceived time of action was shifted more in trials in which the tone actually occurred, compared to trials where it was omitted. Thus, for the patients with schizophrenia, the tone triggered a retrospective reconstruction of the experience of the action. b| To assess prospective components of sense of agency, the perceived time of action was compared between two blocks of trials, which differed in the probability that the action would cause a tone. Only the trials in which no tone actually occurred are shown. In healthy volunteers, the perceived time of action showed stronger shifts when a tone was highly likely to occur than when the tone was less likely to occur. Thus, the strength of predictions about the tone influenced the experience of action. This prospective aspect of sense of agency was absent in patients with schizophrenia. These results suggested that sense of agency in healthy volunteers is based on predictions that use knowledge about action-outcome relations, whereas sense of agency in patients with schizophrenia is based on retrospective reconstruction. Figure adapted, with permission, from [89].

Fig 5 Coercion reduces sense of agency

a| A schematic illustrating an experiment testing the effects of coercion on sense of agency[92]. The participant was asked to estimate the time interval between a button press and a tone. In the coercive condition the experimenter told the agent which key they had to press before each trial. One key would deliver a painful electric shock to a 'victim' at the same time as the tone, and also bring the agent a financial reward. Another key delivered no shock to the victim, and earned the agent no money. In a free-choice condition, the agent could choose between the same two keys, and the experimenter gave no instructions. The agent's estimates of intervals between action and tone were significantly greater when they were coerced into giving the victim a shock, than when they freely chose to do so, suggesting a reduced sense of agency under coercion. b| Event-related potentials (ERPs) evoked by the tone. Coercive instructions reduced the amplitude of auditory N1 component of the ERP, suggesting that the neural processing of action outcomes was reduced in the coercive condition (dark blue) compared to a free-choice condition (light blue). This effect was seen both when the action lead to a painful shock, and when it did not. Panel b reproduced, with permission, from [92].

Glossary terms

Instrumental control: The capacity to initiate an action, and thus bring about an intended change in the environment.

Mirror self-recognition*:* The capacity to recognize a visual percept as being related to one's own body. This has traditionally been assessed by a test in which a colored mark is placed on a body location , such as the forehead, and subsequently viewed via a mirror. Only if the animal recognizes the body in the mirror as its own will it try to remove the mark.

Volition*:* The process of preparing, initiating and executing an action under one's own control. Traditionally, the hallmark of volitional action is that the agent 'could have done otherwise', implying that the action was not directly caused by the current stimulus environment.

Transcranial Magnetic Stimulation (TMS): A technique in which a strong magnetic field is applied to the scalp to influence neural activity in a cortical area beneath. If ongoing cognitive performance is impaired, the affected cortical area is assumed to be necessary for the task.

Transcranial Direct Current Stimulation (TDCS): A non-invasive brain stimulation technique in which a small current passes between electrodes positioned on the scalp. Anodal stimulation is thought to increase excitability of the underlying cortex, while cathodal stimulation may reduce excitability.

Efference copy: A copy of the outgoing (efferent) motor command from the brain to the muscles. An efference copy, in conjunction with a forward model, can be used to predict the sensory consequences of action.

Instrumental actions: Actions that produce a direct or indirect consequence on an animal's external environment. The transformation of the environment is the goal of the action.

Prediction error: The difference between the actual outcome of an action and the predicted outcome. Neural signaling of prediction error can be used to adjust and improve performance, and also to learn how to improve future predictions.

Event-related potential (ERP): An electrical potential that is generated in the brain as a consequence of neuronal activity becoming synchronized by the external stimulus. ERPs are recorded by averaging EEG recorded at the scalp and time-locked to a stimulus. ERPs consist of precisely timed sequences of waves or 'components', which may each reflect a specific cognitive process in the brain.

Online Only

Author Biography

Patrick Haggard obtained his PhD from Cambridge University in 1991, with a thesis on coordination of human manual movement. He has been at University College London since 1995, where he leads a research group on Action and Body at the Institute of Cognitive Neuroscience. He has published several papers on the neural and cognitive processes of voluntary action, with a particular focus on understanding the experience of agency in health and disease.

Key Points

Sense of agency refers to the feeling of controlling one's own actions, and through them, events in the external world.

Sense of agency can be measured in experimental settings by asking participants to explicitly judge whether their action caused an outcome event, or by using implicit measures, such as the compression of perceived time between action and outcome.

Current models of motor control propose that sense of agency is established retrospectively, by comparing delayed sensory feedback about actions and their consequences, against the feedback predicted by an internal model. Connectivity between the frontal areas that develop motor plans for voluntary action, and the parietal areas that monitor outcomes, plays a key role in computing sense of agency.

Processes in the frontal cortex occurring prior to the initiation of action also contribute to sense of agency. For example, selecting which of a number of alternative actions to make can increase sense of agency over the subsequent outcome. These frontal contributions to agency operate prospectively, and underlie the metacognitive experience of one's own voluntary action.

Several neuropsychiatric disorders involve distorted or unreliable sense of agency. This suggests that successful computation of agency by the brain is a key element of normal consciousness and mental health.

Many key features of modern human societies, such as social responsibility, or use of advanced technologies, are based on the brain's ability to compute agency correctly, even in complex interactions.