

75 Intention and Agency

PATRICK HAGGARD

ABSTRACT Cognitive neuroscience makes an important distinction between actions that are triggered by a specific stimulus in the external world (i.e., reactions) and actions that are relatively independent of any individual stimulus. These internally-generated actions instead result in a combination of motivation to act and an internal decision about what action to make and when to make it. Some internally-generated actions are accompanied by a characteristic conscious experience of “willing” or intending to move, which is clearly linked to the ability to initiate the action. This chapter focuses on the brain mechanisms and computations underlying intentional action. The medial frontal cortex plays a key role in generating intentional actions and, in conjunction with the parietal cortex, in the subjective experience of intention. The second part of the chapter focuses on the sense of agency—the experience that our actions aim at external outcomes and are the causes of those outcomes. The same frontal brain circuits that initiate intentional action are also responsible for linking actions to outcomes. Some evidence suggests the action-outcome link is a prospective prediction of outcomes within action-related circuits, rather than a retrospective causal inference within outcome-related circuits. The chapter ends with a brief discussion of the implications of cognitive neuroscience for social and moral responsibility.

Intentions are mental states that represent one’s future actions. However, not all actions are intentional. The key tasks of the cognitive neuroscience of human intentional action are to identify the specific features of intentional actions as opposed to other kinds of actions, to investigate the specific mechanisms of intentional action in the brain, and to consider the implications of these mechanisms for human cognition in general.

The cognitive neurosciences’ interest in intentional action emerges from two sources: the classic philosophical interest in free will and the increasing mechanistic understanding of the motor circuits in the brain. Healthy human adults generally have the experience that they can choose their actions for themselves on at least some occasions. This makes them responsible for an action and its outcomes. However, the neurosciences view actions as events in the motor system, caused by neural mechanisms in cortical and subcortical cognitive and motor areas. The first view is person-centered and subjective, while the second is system-centered and mechanistic. Reconciling these two very different views of human action represents an important frontier in our scientific understanding of ourselves. Moreover, scientific investigation of intention and agency may be

relevant for the diagnosis and treatment of the many neuropsychiatric disorders that affect capacity for voluntary action. Third, and perhaps most importantly, neuroscientific findings may have important social implications, because human society depends on a concept of individual autonomy and responsibility for action.

What is intentional action?

Ludwig 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?” (Wittgenstein, 1953, §621). We all know that intentionally lifting one’s arm and having one’s arm lifted feel quite different, but cognitive psychology has nevertheless struggled to provide a good definition of intentional actions, where there is a direct and immediate correspondence between stimulus and action. Intentional actions are not driven by any obvious or immediate external stimulus. For example, my leg may kick forward because of a sharp tap on the tendon just below the kneecap (a reflex), or because I just decide to kick my leg forward. In essence, this approach defines intentional actions as internally-generated, rather than externally-triggered (Passingham, Bengtsson, & Lau, 2010). This definition has dominated recent thinking about intentional action, and has both advantages and disadvantages. The main disadvantage is that the definition says what intentional actions are not, but does not say what they are in a positive sense. In particular, the causes of intentional action remain largely undefined. On the other hand, this definition gives a clear operational definition of intentional action that can be used in experimental designs. Indeed, many studies have compared the brain mechanisms for internally-generated versus externally-triggered actions. These studies are reviewed in the next section.

Several studies have compared the brain processes for internally-generated versus externally-triggered actions. The design of these studies aims to make the physical movement, and any physical stimulation, balanced across experimental conditions. Differences between conditions can then be linked to the different causes of action. For example, an early positron emission tomography study (Deiber et al., 1991) asked

participants to move a joystick at regular intervals. The direction in which to move the joystick was either instructed by an external signal or freely chosen by the participants. The results showed increased neural activity in the supplementary motor cortex when actions were selected by internal cues, compared to external cues. A subsequent study (Deiber, Ibañez, Sadato, & Hallett, 1996) replicated this difference even when the external cues allowed the same degree of motor preparation as free selection of action, and showed that the centers specifically involved with internal generation of action were located more anteriorly within the supplementary motor cortex, in the region subsequently labelled the presupplementary motor area (preSMA).

More recent functional MRI studies have contrasted free selection with external guidance for several different dimensions of action decision. In the studies of Kriehoff, Brass, Prinz, and Waszak (2009) and Mueller, Brass, Waszak, and Prinz (2007), participants selected which of two actions to make on the basis of external cues or their own free choice. In another condition, the action itself was fixed, but participants could decide when to make it, again based on either external cues or free choice. Freely selecting which action to make, as opposed to being instructed which action to make, activated the rostral cingulate zone (Mueller et al., 2007). A subsequent study replicated this finding, and also showed that freely selecting when to act activated a slightly more posterior area on the surface of the superior frontal gyrus (Kriehoff et al., 2009).

The above findings fit with a general account of a fundamental gradient in the specialization of frontal cortex, between internal guidance of action on the medial aspect and external triggering on the lateral aspect. Medial and lateral frontal regions have different connectivity: the lateral frontal cortex receives from the inferior parietal regions that form the extension of the dorsal visual stream, while the medial frontal regions receive from limbic and superior parietal regions (Averbeck, Battaglia-Mayer, Guglielmo, & Caminiti, 2009).

Perhaps the strongest evidence for a concept of internally guided action comes from lesion studies. Monkeys learned to perform an arbitrary movement (raising the arm) to receive food. When the medial part of the premotor cortex was surgically ablated, the frequency of this action was sharply reduced, although the animals could make the action quite normally in response to a tone (Thaler, Chen, Nixon, Stern, & Passingham, 1995). Recordings from single units in the medial and lateral premotor cortex confirmed this preference. Monkeys learned a sequence of three successive movements. They then performed these movements either on the

basis of internally-stored information or in a condition where an external visual cue specified each movement in turn. The majority of cells in the medial frontal areas, notably the supplementary motor area (SMA), showed a preference for internally-generated movement, while the majority of cells located more laterally in the premotor cortex showed a preference for externally-cued movement (Halsband, Matsuzaka, & Tanji, 1994).

Intentional action as internal generation of motor information

Internally-guided and externally-triggered actions are opposite ends of a cognitive continuum. Most everyday actions probably lie somewhere in between, and involve a mixture of one's own decisions and responses to the external environment. Many readers of this book will have had the experience of participating in a psychology experiment. You typically decide for yourself to participate (signing the consent form being the external marker of this decision), but then you may respond to the stimuli the experimenter shows you. It is hard to say exactly which parts of the activity are intentional, as opposed to stimulus-triggered, and what specific factor makes them intentional. In fact, intentional actions are characterized by the presence of several different factors, none of which may be necessary, but which may be jointly sufficient. We describe some of these below.

First, intentional actions result from internal *decisions* about what to do, when to do it, and even whether to do it at all. Mike Shadlen has proposed the term “freedom from immediacy” for this core property of human and some nonhuman action selection (Shadlen & Gold, 2004). Clearly, actions can be more or less constrained by the environment, so the degree of internally-generated decision making must be graded.

Second, intentional actions normally have a distinctive *motivation*: they are goal-directed or aim at producing a particular outcome. This indicates a tight linkage between intentional action and neural reward systems. Humans and animals internally generate actions that produce appropriate reinforcement, even when there is no external stimulus cuing the action. Somatic states such as hunger and thirst may form the basis of intentional action (Damasio, 2000). However, the human experimental literature has struggled to include this motivational aspect in experimental designs. Given a transparent choice between more rewarding and less rewarding actions, or between more and less moral outcomes, people generally choose the most desirable action (Moretto, Walsh, & Haggard, 2011). For example, if I offer someone a \$10 bill or a \$20 bill, they will probably reach for the \$20 bill—but this seems more like a

direct response to the \$20 stimulus than an internally-generated action. Because of this ambiguity, many human experiments used arbitrary, valueless action decisions. Such experiments clearly cannot capture how everyday intentional actions follow from goals and reasons, but they may be able to study the process of generating intentional action.

A third critical element of intention is the *timing of initiation* of action itself. The combination of a motivation to act and a decision about which action to make is not always sufficient to trigger action. For example, action may be physically prevented by external constraint, or not appropriate given current circumstances, competing actions, or competing goals. For these reasons, the brain needs a mechanism to maintain intentions and release them only at the appropriate time: another form of “freedom from immediacy.” The temporal aspect of intention has received more experimental attention than any other, and forms the topic of the next section.

The combination of decision, motivation, and timing of initiation gives intentional actions a level of flexibility and complexity that stimulus-driven actions lack. The traditional definition of intentional action can sometimes be misinterpreted as implying an action with no apparent cause—something that is clearly nonsensical. Perhaps it is better to think of internally generated actions as responses where a combination of several factors jointly cause action (Schüür & Haggard, 2011).

Mental chronometry of intentions

Mental chronometry has been a rich line of investigation in many areas of psychology, and no more so than in intentional action. Recent computational models (see figure 75.1) provide a useful framework for considering intentions and time.

In these models (see chapter 39), actions begin with specification of a goal. A “planner,” or inverse model, converts the goal into a set of motor commands to achieve the goal. These commands are sent from the brain to the spinal cord and the muscles, causing physical movement and finally bringing about the goal state. A copy of the motor command is sent to an internal forward model, which predicts the effects of the motor command. Therefore, the system potentially has access to several different kinds of information about a current action, each relating to a different time: a description of the future or goal state, an estimate of the current state, and a delayed description based on feedback from the body and the external world.

This model of intentional action begins with goals, but the concept of action goal has been viewed in two

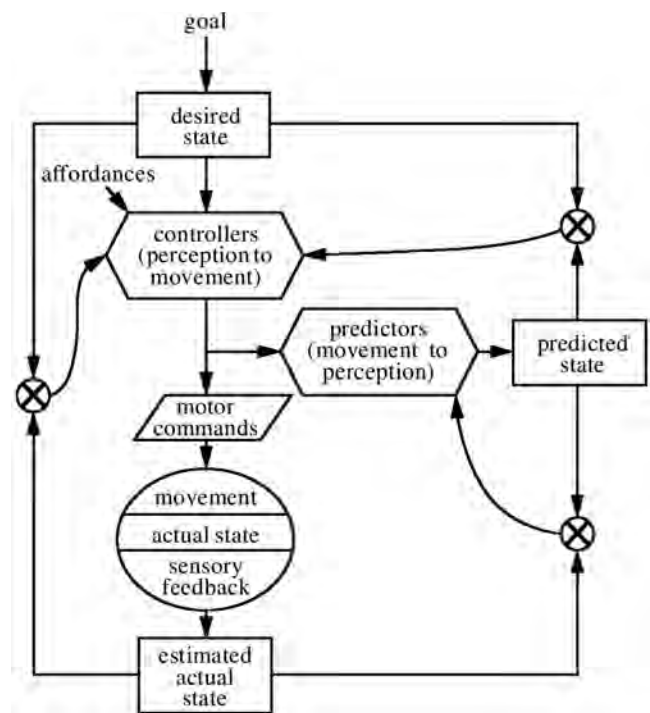


FIGURE 75.1 Computational motor control model for goal-directed action. (Reprinted with permission from Frith et al., 2000. Copyright © 2000, The Royal Society.)

quite different ways within cognitive neuroscience. In cognitive psychology, intentional action goals are often considered as prospective memories to perform a desired action at a later, appropriate time. In sensorimotor control, the goal is a state of the body, such as getting the hand to a particular target position in space. The main difference between the two approaches is how close the person is to the time of action. Prospective memory involves long-term or prior intentions: other events and other actions may occur before the intended action. Sensorimotor control models, in contrast, begin with the immediate current intention or goal and are assumed to operate over the short term. They therefore deal with intentions-in-action rather than long-range intentions (Searle, 1983). An external triggering event, or an internal representation of elapsed time, is required to retrieve a relevant prior intention from long-term memory and transform it to an intention-in-action: this process marks the trigger or decision to make the action *now*. This progression toward triggering action is associated with a gradient across the prefrontal cortex: more anterior regions deal with more abstract, longer-range intentions, while more posterior regions deal with specific action details (Koechlin & Summerfield, 2007; Pacherie, 2008).

Neuroimaging studies identify the lateral part of Brodmann’s area (BA) 10 in the prefrontal cortex as a

key area for *maintaining* intentions (Burgess, Gonen-Yaacovi, & Volle, 2011). However, fewer studies have investigated how a stored intention is transformed into an action. One recent study used multivariate pattern analysis of fMRI data to identify the regions of the brain from which it was possible to decode which of two tasks a person would perform, and also when they would switch to the new task (Momennejad & Haynes, 2012). Two distinct regions of the medial prefrontal cortex carried information relevant to timing of intention. The lateral prefrontal cortex in both hemispheres *maintained* information about the forthcoming switch. At the time of retrieval, and at switching to the new task, information about timing was present more medially in BA10 as well as bilaterally in the SMA. While this study cannot causally identify the trigger signal that means “act now,” it confirms a lateral-to-medial as well as an anterior-posterior gradient in prefrontal cortex in transforming long-range intentions into current actions.

Conscious intention and brain activity

A key question in cognitive neuroscience of intentional action has been the relation between conscious intention and the brain processes that trigger action. The processes of action selection, motivation, and temporal flexibility that characterize intentional action are all associated with specific conscious experiences, of choosing, of urge, and of commitment to act. I have the experience that “I” am the cause of my own actions, while the conscious experiences associated with reflex actions and with habitual actions are quite different. In reflex action, we are generally aware of the triggering stimulus and are sometimes surprised by our own motor responses. In habitual or routine actions such as walking or typing, we experience a background buzz of ongoing activity, without clear awareness of each individual motor command.

In the last decades, several neuroscientists have investigated the neural basis of the characteristic conscious experiences accompanying intentional action. Sometimes this work has been used to consider metaphysical questions, such as whether humans have free will or not. Here we limit ourselves to a “natural history” approach to intention, asking what conscious experiences accompany intentional action and what mechanisms produce those experiences. Whether this data constitutes evidence of “free will” is a philosophical rather than a neuroscientific point, and the subject of many excellent books (e.g., Kane, 2005).

Inquiry in this field began with the famous “Libet experiment” (Libet, Gleason, Wright, & Pearl, 1983). Libet asked participants to make a simple voluntary

action (a movement of the wrist) at a time of their own choosing. EEG was recorded from the scalp throughout the experiment. In addition, participants watched a spot that rotated continuously on a clock face. After each action, participants indicated where the spot had been when they first “felt the urge to move.” Libet’s finding was simple, but striking (figure 75.2). People experienced the urge to move on average 206 msec before the physical onset of muscle contraction. However, the preparatory activity in the brain, or “readiness potential” (Kornhuber & Deecke, 1965) was present as much as 1 sec prior to this point. This temporal sequence, in which brain activity comes first and conscious awareness of impending movement comes later, seems to rule out the Cartesian dualist concept of conscious free will, according to which the conscious controls movements of the body via the intermediate mechanisms of the brain. Libet concluded that voluntary acts are initiated by unconscious brain mechanisms, rather than by conscious intention.

This result has attracted considerable attention in philosophy, and in the public imagination. The experimental method suffers from several drawbacks. For example, simple voluntary actions often do not produce a single, discrete, vivid experience of “urge,” or intention, or decision. Estimating the time of conscious intention is more difficult than estimating the time of an auditory onset or visual onset. Some have questioned whether such judgments make sense at all. Nevertheless, the basic result seems to replicate: people can report an experience a few hundred milliseconds prior to voluntary action. It is less clear what gives rise to this experience, and what the content of the experience is.

From a neuroscientific point of view, the Libet experiment can seem almost trivial: what could conscious intention be other than a product of brain activity?

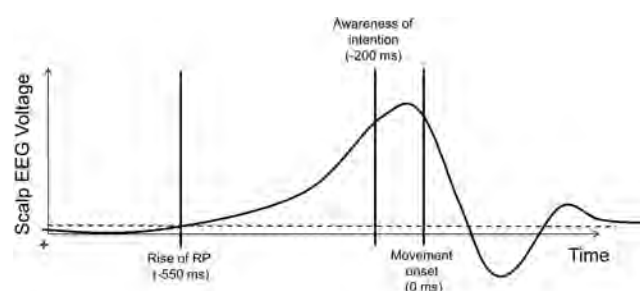


FIGURE 75.2 The Libet experiment. The buildup of a “readiness potential” preceding voluntary action occurs several hundred milliseconds before action onset. In contrast, the subjective experience of intention is reported to occur around 200 msec before action onset. Libet argued from the temporal order of these events that conscious intention could not be the trigger for action.

However, since numerous brain processes contribute to preparation and control of actions, an important scientific question is to identify those that underlie the experience of intention. Fried and colleagues (Fried, Mukamel, & Kreiman, 2011) recorded from 1,019 individual neurons in the brains of patients undergoing preoperative monitoring before epilepsy surgery, while the patients performed the Libet experiment. They found several neurons in the medial frontal cortex that progressively modulated their firing rate before the patients' average reported time of conscious intention. Interestingly, these neurons could either increase or decrease their firing rates as the action onset approached, suggesting that medial frontal areas may initiate action by a balance of excitatory and inhibitory drives onto motor-execution areas. We shall return to this point later. They also showed that they could accurately predict the time of the conscious intention to move from a population of a few hundred such medial frontal neurons. Neurons in other brain areas, such as the temporal lobe, showed much less association with conscious intention.

“Negative volition” and inhibition of action

An interesting recent debate has focused on decisions to inhibit intentional actions. Behavioral sciences have often struggled to study inhibition of action, largely because it produces no measurable behavioral output. As a result, the contribution of inhibitory processes to intentional action has generally been underestimated. Nevertheless, classical neuropsychological studies showed that medial frontal lobe damage could cause a form of excessive, compulsive action, in which the patient's actions would be triggered automatically by objects that happened to be present in the environment, even when no particular reason or desire to perform the action could be detected. These were termed utilization behaviors (Lhermitte, 1983). This finding suggests that one normal function of the healthy medial frontal cortex may be to inhibit those actions that we currently do not wish to make, and not simply to initiate those actions we do wish to make. A dramatic example comes from the neurological syndrome of anarchic hand. This unilateral form of utilization behavior sometimes follows a medial frontal and callosal lesion. Della Sala and colleagues (Della Sala, Marchetti, & Spinnler, 1991) report the case of a patient whose anarchic right hand reached out involuntarily to grab a hot beverage immediately after the patient announced that she had decided to wait for the drink to cool before taking it. This compelling example suggests that the maintenance, and then eventual release, of inhibition

may be part of the normal regulatory process that allows us to make an action only when it is appropriate, and not at other times when it would be inappropriate. Recent work in healthy humans (van den Wildenberg et al., 2010) and monkeys (Wardak, 2011) confirms that deciding *whether* to act and deciding *when* to act are closely linked. Even in a simple reaction task, cognitive motor areas maintain a tonic inhibitory influence on the primary motor cortex until just before the time of the expected “go” signal (Duque et al., 2010).

Intentional inhibition has sometimes been viewed dualistically, as a form of “conscious veto” underlying self-control (Libet, 2005). However, the neuropsychological cases described above show that a specific frontal brain mechanism is responsible for inhibiting actions that might, in principle, be executed. Neuroimaging studies of inhibition broadly agree with this view. The classic experimental paradigms for studying inhibition generally involve an external *stop* or *no go* signal—yet in everyday life, healthy adults are expected to inhibit inappropriate actions endogenously, without any explicit instruction (Aron, 2011; Filevich, Kühn, & Haggard, 2012). Therefore, intentional paradigms such as the Libet experiment have also been used to investigate inhibition (Brass & Haggard, 2007). Participants were asked to prepare voluntary key-press actions, but then to cancel their action at the last possible moment on a freely chosen subset of trials. These intentional inhibition trials present an unusual experimental challenge, because the input is not directly controlled by the experimenter's instructions, nor is there any behavioral output to measure! Crucially, however, participants indicated the time at which they experienced willing the action using a rotating clock hand, even if the action was subsequently canceled. This subjective marker allowed the authors to investigate brain activity time-locked to intending actions and then inhibiting them. A region of BA 9 in the dorsomedial prefrontal cortex was found to be activated in inhibition trials, but deactivated in action trials. A later study, using a different experimental paradigm, additionally showed that this area had a strong effective connectivity with the preSMA during inhibition trials, relative to action trials (Kühn, Haggard, & Brass, 2009).

Based on this evidence, it was suggested that this dorsomedial frontal area may express the intention to inhibit actions, and may do so by exerting control over action-preparation circuits elsewhere in the frontal cortex. Interestingly, a similar prefrontal region, with similar connectivity to preSMA, was activated when participants intentionally decided to resist feeling the emotion suggested by unpleasant visual stimuli (Kühn, Haggard, & Brass, 2013). This region may therefore

form an important part of a general brain circuit underlying inhibitory self-control. Consistent with this view, hypoactivity of the medial prefrontal cortex during inhibition tasks has regularly been reported in studies of ADHD (Rubia et al., 1999; Smith et al., 2006).

Sense of agency

The sense of agency refers to the feeling that one controls one's own actions and, through them, events in the external world. Sense of agency thus refers to the experience and mental representation of the relation between one's own intentional actions and their external sensory consequences. It is difficult to imagine a more fundamental distinctive feature of human mental life: all human endeavors, technologies, and transformations of our environment are ultimately based on being aware of the consequences of our actions. Moreover, sense of agency plays a crucial social role: agents can only be held responsible for the consequences of their actions if there is general agreement on who performed the action. For example, many legal systems allow a defense based on reduced or absent sense of agency.

Animals can learn and perform instrumental or goal-directed actions very readily. A rat that presses a lever for food perhaps experiences a kind of agency. However, the range and sophistication of human instrumental action is clearly much wider. Agency is best treated as a psychophysical problem: how do people represent and perceive the relation between intentional actions and outcomes? Like all psychophysical investigations, we begin with physical reality: these are the *facts* of agency, whether the agent performed the action or not. Next, we can consider the agent's awareness deriving from these facts. The term "sense of agency" refers to the normal awareness of initiating one's voluntary actions, and thus controlling their immediate outcomes. For many habitual actions, the sense of agency is thin, but it remains present as a background "buzz," accompanying normal mental life (Synofzik, Vosgerau, & Newen, 2008). Sense of agency moves to the foreground of consciousness during significant action decisions or when action outcomes are important.

Most scientific studies have not tackled the conscious experience of agency directly, but have asked participants to make binary *judgments* about whether they did or did not cause a particular action outcome. These studies have identified ambiguous situations in which judgments of agency may be incorrect, and have also identified the physical clues used for making such judgments. Farrer and colleagues (2008) instructed participants to move a set of wooden pegs into holes while in

an fMRI scanner. Participants viewed a video image of their own action with a variable delay. In a first condition, participants simply detected whether the image was delayed or not, while in a second condition, participants judged whether the video they saw showed their own action or not—that is, they made judgments of agency. The behavioral results from the first condition showed that delays in visual feedback became detectable only if they lasted 200–300 msec or more. The results from the second condition showed that this time window produced an ambiguous sense of agency, with participants sometimes judging the observed action as their own and sometimes not. Comparing positive and negative judgments of agency for these intermediate delays did not identify any brain area coding for positive judgments of agency. However, the angular gyrus showed a stronger activation for trials where participants denied agency, compared to trials where they accepted it. Interestingly, neuroimaging results from the first, delay-detection condition also showed that angular gyrus activation increased with the video delay. This study confirms and extends previous reports of angular gyrus involvement in agency judgement (Sirigu, Daprati, Pradat-Diehl, Franck, & Jeannerod, 1999). It further suggests that the sense of being in control of one's actions is a default mode of the brain, not specifically coded as an individual state. Second, timing provides an important cue for the function of this circuit.

The computational motor control model (figure 75.1) has been used to explain the sense of agency. If the comparison between the state of the limb estimated by an internal model matches that reported by sensory signals, the current sensory input is attributed to one's own action. This causes a sense of agency over the current sensory input. In the model, as in the human neuroimaging literature, agency is not defined by any positive signal but only by an absence of errors. Moreover, the model insists that sense of agency is necessarily retrospective: the brain waits for delayed sensory feedback before the computations that lead to sense of agency can begin. This link between feedback delay and agency recalls the pattern of results of the neuroimaging results described above.

The same model has been used to explain why the consequences of self-produced action are often not perceived: conscious awareness is associated not with actual sensory feedback but with the element of sensory input that is not predicted by one's own motor command (Blakemore, Wolpert, & Frith, 2000; Gentsch & Schütz-Bosbach, 2011). However, it may seem paradoxical to use the same model both to account for the experience of controlling external events and to account for the

relative imperception of external events caused by one’s own agency.

A recent study resolved some of these uncertainties with a novel interpretation of the role of the angular gyrus in sense of agency. Chambon and colleagues (Chambon, Wenke, Fleming, Prinz, & Haggard, 2013; see figure 75.3) used subliminal arrow stimuli to prime participants’ actions, so that selecting the appropriate action in response to a left or right target arrow was either facilitated or inhibited. After the participants responded, one of several color patches appeared on the screen after a short delay. Participants reported how much control they felt they had over the color appearing on the screen. The results showed a higher sense of control over action outcomes when participants’ responses had been facilitated by subliminal priming, relative to when they had been impaired. Interestingly, the angular gyrus showed increasing activation as the

participants’ feeling of control reduced, but only for incompatibly primed trials. Crucially, the primes in this experiment did not predict the color patches that appeared after each action. Therefore, the influence of priming on sense of agency was interpreted as a prospective fluency effect: people feel a stronger sense of control when it is easy to *select* which action to take, irrespective of the statistical relation between their action and outcome. People mistake the ease of choosing what to do for actually achieving something. Interestingly, the angular gyrus showed increasing levels of activation as the subjective sense of control over the color path decreased, but only for incompatible trials. This result confirms the negative coding for nonagency in the angular gyrus. However, this code may be prospectively based on what we think we may achieve, and not just retrospectively based on what we have actually achieved.

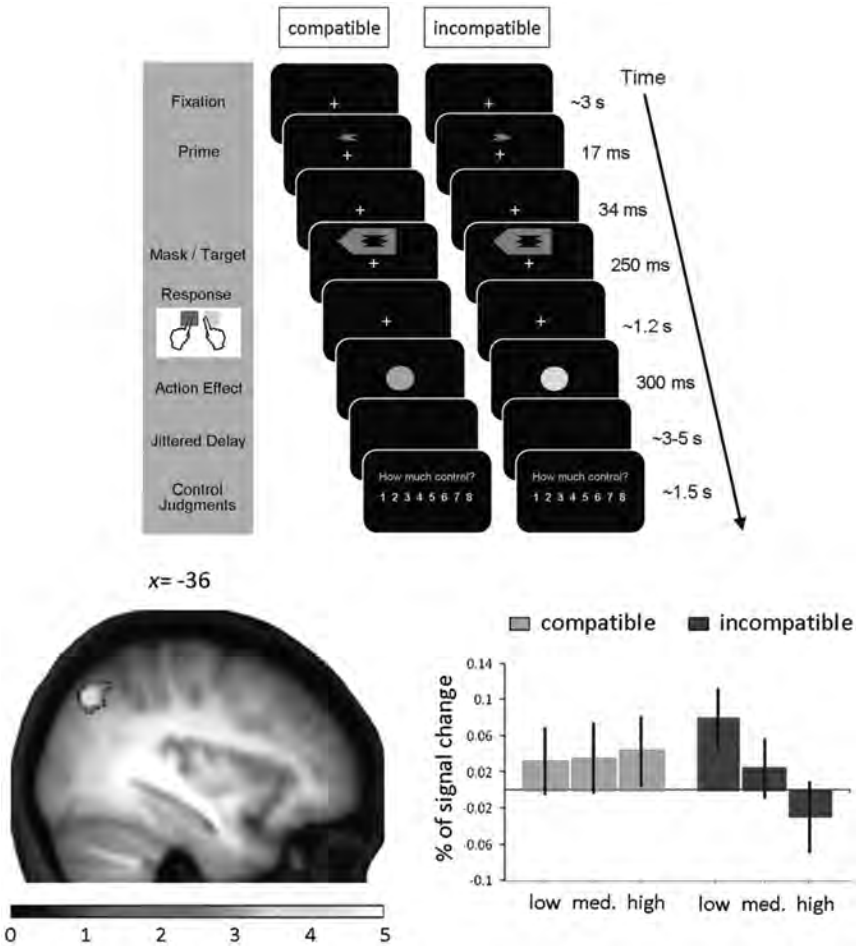


FIGURE 75.3 Subliminal priming of actions influences sense of agency *prospectively*. *Upper panel:* subliminal primes facilitate or impair actions in response to target arrow stimuli. Responses are followed by appearance of a color patch, and participants judge how much control they have over the color patch. *Lower*

panel: activation of the angular gyrus associated with the target-response event varies negatively with perceived level of control, but only following incompatible priming. (Reproduced from Chambon et al., 2013, by permission of Oxford University Press.) (See color plate 67.)

Low-level agency perception

Several studies have shown that human judgments of agency may be biased. In social situations, people tend to think they caused events that were in fact caused by the actions of others (Wegner, 2003; Wegner & Wheatley, 1999). The bias to overestimate one's own agency in the case of simple sensorimotor events may form an instance of a more general tendency to believe in one's own self-efficacy (Bandura, 2001). However, it remains unclear whether these biases actually alter low-level perception of agency, or whether they are merely response biases. Investigating this question requires an implicit measure of agency. One candidate is the "intentional binding" effect. In the original demonstration (Haggard, Clark, & Kalogeras, 2002), participants used the clock method developed by Libet and colleagues to indicate the time at which they made a voluntary key-press action. In one block of trials, each action was followed after a short delay by an auditory tone, while in another block of trials no tone occurred. The perceived time of the actions was shifted later, toward the subsequent tone, in the block where tones occurred. In the same fashion, the perceived time of the tone was compared between a baseline block where participants simply heard a tone occurring at random intervals, without making any action, and an experimental block in which participants caused the tone through their own voluntary action. The perceived time of the tone was shifted earlier, toward the action that caused it, compared to the baseline block. Taken together, the action binding and the tone binding implied a perceptual compression of the time interval between action and outcome. Crucially, neither of these effects occurred when the voluntary key-press action was replaced by an involuntary twitch of the same muscles induced by transcranial magnetic stimulation (Haggard, Clark, & Kalogeras, 2002; see figure 75.4), or by a second tone. On this basis, it was suggested that intentional binding may reflect the temporal association between intentional actions and their outcomes. It may therefore provide an implicit marker of sense of agency.

The intentional binding effect involves independent measures of action binding and tone binding. Therefore, intentional binding allows sense of agency to be broken down into distinct components, in a way that judgment of agency attribution does not. For example, by varying the probability of a tone following an action, Moore and Haggard (2008) showed that action binding involved two distinct processes, one prospective and one retrospective. In a block where only 50% of actions were followed by tones, action binding was stronger on

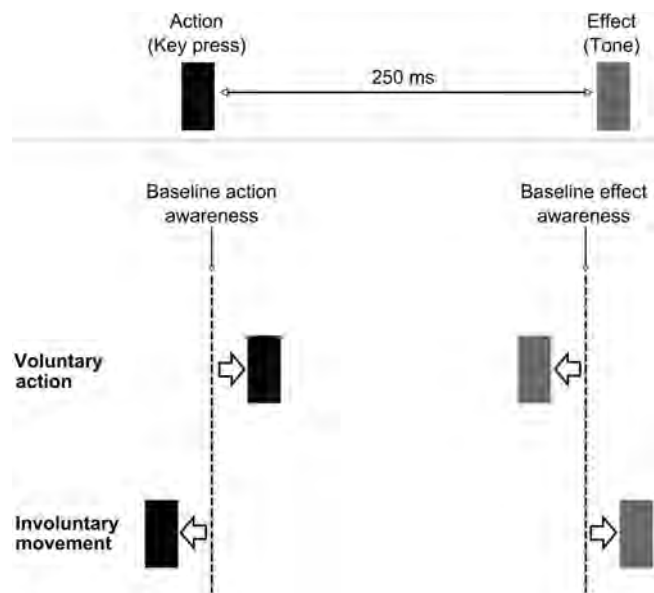


FIGURE 75.4 The intentional binding effect. (Reproduced from Moore et al., 2013, with permission from Elsevier.)

trials where the tone occurred than on trials where it did not, suggesting a change in the perceived time of the action triggered retrospectively by the ensuing tone. In a block where most actions (75%) were followed by tones, action binding was again observed on trials where the tone did occur, but also on trials without any tone. This pattern of results suggests that the action was shifted prospectively toward the tone, in anticipation of its likely occurrence. Voss and colleagues used a similar design to compare the predictive and retrospective processes of action binding between a group of psychotic patients and age-matched healthy controls (Voss et al., 2010). They found similar levels of binding in both groups, but for very different reasons. In the healthy controls, action binding was dominated by the prospective component, suggesting it depended on predicting the likely outcomes of actions. In the patients, action binding was dominated by the retrospective component. Disorganization of thought and agency are cardinal symptoms of schizophrenic psychosis. This result suggests that even low-level experience of simple sensorimotor events in psychosis depends on opportunistic conjunctions of events, rather than learned regularities of instrumental action.

Studies of the neural basis of intentional binding have suggested a key role for several frontal lobe circuits. Importantly, these show considerable overlap with the areas involved in generating intentional action. They are also quite distinct from the parietal areas generally identified with explicit judgments of agency.

This evidence can only be reviewed briefly here. First, intentional binding effects were reduced after inactivation of the preSMA caused by repetitive transcranial magnetic stimulation (Moore, Ruge, Wenke, Rothwell, & Haggard, 2010). Second, Parkinsonian patients showed significantly stronger intentional binding after taking their dopaminergic medication than immediately before, suggesting that dopaminergic drive in the basal-ganglia-thalamocortical loop strongly contributes to sense of agency (Moore, Schneider, et al., 2010). Finally, an fMRI study showed that activation of the caudal SMA correlated more strongly with the perceived interval between voluntary actions and tones than between passive movements and identical tones (Kühn, Brass, & Haggard, 2012).

Context and belief in the conscious experience of intentional action

Taken together, these results suggest that the processes of preparing and initiating voluntary actions in the frontal lobes also involve a prediction of the anticipated consequences of action, much as the computational motor control model suggests. Importantly, the experience of agency, or causing the effects of one's action, appears to derive from the same circuits that develop action itself. Thus, the sense of agency may not simply be an inference or confabulation of authorship, but a measurable internal signal within the motor system. There is currently a lively controversy regarding the division of labor between the frontal and parietal lobes in both intention and agency. The conscious urge to move that was previously identified with medial frontal cortex was also recently reported after direct electrical stimulation of parietal cortex (Desmurget et al., 2009), while patients with parietal lesions had a delayed experience of intention to move (Sirigu et al., 2004). The involvement of parietal cortex in agency is also clear, since the angular gyrus is routinely activated in judgments of (non)agency, as we have seen. However, frontal activations did correlate with implicit markers of sense of agency, whereas parietal activations did not (Kühn et al., 2012). Therefore, one interim hypothesis suggests that the frontal lobe houses the core machinery for voluntary action, including the prospective construction of experience of goal-directed action and the low-level experience of agency. The parietal lobes, in contrast, would monitor the motor instructions generated frontally. By comparing these instructions with later sensory inputs, a parietal comparator could detect nonagency and contribute to explicit judgments of agency.

Responsibility for action

All human societies appear to hold individuals responsible for their actions. In many systems of law, theories of responsibility are explicitly dualist. The agents' capacity for rational thought, their conscious choice over their actions, and their freedom to do otherwise—all make them responsible for what they do. In dramatic contrast, neuroscience ultimately views intentional actions as consequences of specific, deterministic electrical and chemical brain processes. Modern cognitive neuroscience rejects dualist notions of brain-independent consciousness, and instead views an individual's choices, actions, and even his or her character as mechanistic functions of his or her brain. For the neuroscientist, the conscious, rational thought is simply an additional brain process, which is presumably just as determined as all others. These opposing views form the basis of a lively debate between law and neuroscience. For example, if psychopathy is a neuropsychiatric disorder, are psychopaths truly responsible for their actions? Can neuroscience help to guide such individuals toward more prosocial behaviors (Anderson & Kiehl, 2013)? Greene and Cohen argue that if neurobiological determinism is true, then retributive justice and punishment are misguided. To the extent that a criminal's brain "made them do it," then there is little point in punishing criminals for their actions (Greene & Cohen, 2004). Of course, the same determinist neuroscientific theories of action offer ready support for a rehabilitative view of justice. Rehabilitation of offenders could, in principle, provide learning experiences that change the brain, and thus behavior.

Humans are social animals, and our brain mechanisms have coevolved with our increasingly large and diverse societies. To participate in such a society requires a brain that is able to follow, or learn to follow, the social and moral codes by which societies work. Equally, societies will generally try to protect themselves against possible harm from individuals who do not follow these codes. The methods of protection are costly, since they involve either costly care for those who are incapable of following such codes or costly punishment for those who are judged capable, but who transgress. For a strict neurobiological determinist, individuals are punished not so much for the action that they performed, but for having the kind of brain they have—since it was their brain that caused their action. The consequences of this view are highly unpalatable and morally unacceptable! However, even a determinist may agree that responsibility and punishment are more acceptable in a society where all individuals have an equal opportunity to learn to follow legal and moral codes of behavior than in a

society where some individuals have little or no such opportunity to learn. Education and culture are society's ways of influencing, or constructing, individual brains that produce appropriate behaviors.

The debate on responsibility often ignores this crucial role of learning: responsibility for action presupposes a brain that is capable of learning codes of behavior, and also a brain that has been trained for action by the appropriate learning experiences. If equal access to learning experiences is absent, socially organized responsibility and punishment could simply be seen as social tools for individuals with one kind of brain to control individuals with "less desirable" brains. Perhaps socially organized responsibility and punishment only make sense in societies that ensure equal access to appropriate learning experiences for all.

ACKNOWLEDGMENTS Preparation of this chapter was supported by an ESRC Professorial Fellowship (ES/J023140/1) and by an ERC Advanced Grant HUMVOL (323943).

REFERENCES

- ANDERSON, N. E., & KIEHL, K. A. (2013). Psychopathy: Developmental perspectives and their implications for treatment. *Restor Neurol Neurosci*, 32(1), 103–117.
- ARON, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biol Psychiatry*, 69(12), e55–68.
- AVERBECK, B. B., BATTAGLIA-MAYER, A., GUGLIELMO, C., & CAMINITI, R. (2009). Statistical analysis of parieto-frontal cognitive-motor networks. *J Neurophysiol*, 102(3), 1911–1920.
- BANDURA, A. (2001). Social cognitive theory: An agentic perspective. *Annu Rev Psychol*, 52, 1–26.
- BLAKEMORE, S. J., WOLPERT, D., & FRITH, C. (2000). Why can't you tickle yourself? *NeuroReport*, 11(11), R11–16.
- BRASS, M., & HAGGARD, P. (2007). To do or not to do: The neural signature of self-control. *J Neurosci*, 27(34), 9141–9145.
- BURGESS, P. W., GONEN-YAACOV, G., & VOLLE, E. (2011). Functional neuroimaging studies of prospective memory: What have we learnt so far? *Neuropsychologia*, 49(8), 2246–2257.
- CHAMBON, V., WENKE, D., FLEMING, S. M., PRINZ, W., & HAGGARD, P. (2013). An online neural substrate for sense of agency. *Cereb Cortex*, 23(5), 1031–1037.
- DAMASIO, A. (2000). *The feeling of what happens: Body, emotion and the making of consciousness*. New York, NY: Vintage.
- DEIBER, M. P., IBAÑEZ, V., SADATO, N., & HALLETT, M. (1996). Cerebral structures participating in motor preparation in humans: A positron emission tomography study. *J Neurophysiol*, 75(1), 233–247.
- DEIBER, M. P., PASSINGHAM, R. E., COLEBATCH, J. G., FRISTON, K. J., NIXON, P. D., & FRACKOWIAK, R. S. (1991). Cortical areas and the selection of movement: A study with positron emission tomography. *Exp Brain Res*, 84(2), 393–402.
- DELLA SALA, S., MARCHETTI, C., & SPINLER, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, 29(11), 1113–1127.
- DESMURGET, M., REILLY, K. T., RICHARD, N., SZATHMARI, A., MOTTOLESE, C., & SIRIGU, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, 324(5928), 811–813.
- DUQUE, J., LEW, D., MAZZOCCHIO, R., OLIVIER, E., & IVRY, R. B. (2010). Evidence for two concurrent inhibitory mechanisms during response preparation. *J Neurosci*, 30(10), 3793–3802.
- FARRER, C., FREY, S. H., VAN HORN, J. D., TUNIK, E., TURK, D., INATI, S., & GRAFTON, S. T. (2008). The angular gyrus computes action awareness representations. *Cereb Cortex*, 18(2), 254–261.
- FILEVICH, E., KÜHN, S., & HAGGARD, P. (2012). Intentional inhibition in human action: The power of "no." *Neurosci Biobehav Rev*, 36(4), 1107–1118.
- FRIED, I., MUKAMEL, R., & KREIMAN, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, 69(3), 548–562.
- FRITH, C. D., BLAKEMORE, S.-J., & WOLPERT, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Res Rev*, 31(2-3), 357–363.
- GENTSCH, A., & SCHÜTZ-BOSBACH, S. (2011). I did it: Unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *J Cogn Neurosci*, 23, 3817–3828.
- GREENE, J., & COHEN, J. (2004). For the law, neuroscience changes nothing and everything. *Philos Trans R Soc Lond B Biol Sci*, 359(1451), 1775–1785.
- HAGGARD, P., CLARK, S., & KALOGERAS, J. (2002). Voluntary action and conscious awareness. *Nat Neurosci*, 5(4), 382–385.
- HALSBAND, U., MATSUZAKA, Y., & TANJI, J. (1994). Neuronal activity in the primate supplementary, pre-supplementary and premotor cortex during externally and internally instructed sequential movements. *Neurosci Res*, 20(2), 149–155.
- KANE, R. (2005). *A contemporary introduction to free will*. New York, NY: Oxford University Press.
- KOECHLIN, E., & SUMMERFIELD, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn Sci*, 11(6), 229–235.
- KORNHUBER, H. H., & DEECKE, L. (1965). Changes in the brain potential in voluntary movements and passive movements in man: Readiness potential and reafferent potentials. *Pflugers Arch Gesamte Physiol Menschen Tiere*, 284, 1–17.
- KRIEGHOFF, V., BRASS, M., PRINZ, W., & WASZAK, F. (2009). Dissociating what and when of intentional actions. *Front Hum Neurosci*, 3, 3.
- KÜHN, S., BRASS, M., & HAGGARD, P. (2012). Feeling in control: Neural correlates of experience of agency. *Cortex*, 49(7), 1935–1942.
- KÜHN, S., HAGGARD, P., & BRASS, M. (2009). Intentional inhibition: How the "veto-area" exerts control. *Hum Brain Mapp*, 30(9), 2834–2843.
- KÜHN, S., HAGGARD, P., & BRASS, M. (2013). Differences between endogenous and exogenous emotion inhibition in the human brain. *Brain Struct Funct* (E-pub ahead of print).
- LHERMITTE, F. (1983). "Utilization behaviour" and its relation to lesions of the frontal lobes. *Brain*, 106(Pt. 2), 237–255.
- LIBET, B. (2005). *Mind time: The temporal factor in consciousness*. Cambridge, MA: Harvard University Press.

- LIBET, B., GLEASON, C. A., WRIGHT, E. W., & PEARL, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106(Pt. 3), 623–642.
- MOMENNEJAD, I., & HAYNES, J.-D. (2012). Human anterior prefrontal cortex encodes the “what” and “when” of future intentions. *NeuroImage*, 61(1), 139–148.
- MOORE, J. W., CAMBRIDGE, V. C., MORGAN, H., GIORLANDO, F., ADAPA, R., & FLETCHER, P. C. (2013). Time, action and psychosis: Using subjective time to investigate the effects of ketamine on sense of agency. *Neuropsychologia*, 51(2), 377–384.
- MOORE, J. W., RUGE, D., WENKE, D., ROTHWELL, J., & HAGGARD, P. (2010). Disrupting the experience of control in the human brain: Pre-supplementary motor area contributes to the sense of agency. *Philos Trans R Soc Lond B Biol Sci*, 277(1693), 2503–2509.
- MOORE, J. W., SCHNEIDER, S. A., SCHWINGENSCHUH, P., MORETTO, G., BHATIA, K. P., & HAGGARD, P. (2010). Dopaminergic medication boosts action-effect binding in Parkinson’s disease. *Neuropsychologia*, 48(4), 1125–1132.
- MOORE, J., & HAGGARD, P. (2008). Awareness of action: Inference and prediction. *Conscious Cogn*, 17(1), 136–144.
- MORETTO, G., WALSH, E., & HAGGARD, P. (2011). Experience of agency and sense of responsibility. *Conscious Cogn*, 20(4), 1847–1854.
- MUELLER, V. A., BRASS, M., WASZAK, F., & PRINZ, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, 37(4), 1354–1361.
- PACHERIE, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179–217.
- PASSINGHAM, R. E., BENGTSSON, S. L., & LAU, H. C. (2010). Medial frontal cortex: From self-generated action to reflection on one’s own performance. *Trends Cogn Sci*, 14(1), 16–21.
- RUBIA, K., OVERMEYER, S., TAYLOR, E., BRAMMER, M., WILLIAMS, S. C., SIMMONS, A., & BULLMORE, E. T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: A study with functional MRI. *Am J Psychiatry*, 156(6), 891–896.
- SCHÜÜR, F., & HAGGARD, P. (2011). What are self-generated actions? *Conscious Cogn*, 20(4), 1697–1704.
- SEARLE, J. R. (1983). *Intentionality*. Cambridge, UK: Cambridge University Press.
- SHADLEN, M. N., & GOLD, J. I. (2004). The neurophysiology of decision-making as a window on cognition. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 1229–1241). Cambridge, MA: MIT Press.
- SIRIGU, A., DAPRATI, E., CIANCIA, S., GIRAUX, P., NIGHOGHOSIAN, N., POSADA, A., & HAGGARD, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nat Neurosci*, 7(1), 80–84.
- SIRIGU, A., DAPRATI, E., PRADAT-DIEHL, P., FRANCK, N., & JEANNEROD, M. (1999). Perception of self-generated movement following left parietal lesion. *Brain*, 122(Pt. 10), 1867–1874.
- SMITH, A. B., TAYLOR, E., BRAMMER, M., TOONE, B., & RUBIA, K. (2006). Task-specific hypoactivation in prefrontal and temporoparietal brain regions during motor inhibition and task switching in medication-naïve children and adolescents with attention deficit hyperactivity disorder. *Am J Psychiatry*, 163(6), 1044–1051.
- SYNOFZIK, M., VOSGERAU, G., & NEWEN, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Conscious Cogn*, 17(1), 219–239.
- THALER, D., CHEN, Y. C., NIXON, P. D., STERN, C. E., & PASSINGHAM, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Exp Brain Res*, 102(3), 445–460.
- VAN DEN WILDENBERG, W. P. M., BURLE, B., VIDAL, F., VAN DER MOLEN, M. W., RIDDERINKHOF, K. R., & HASBROUCQ, T. (2010). Mechanisms and dynamics of cortical motor inhibition in the stop-signal paradigm: A TMS study. *J Cogn Neurosci*, 22(2), 225–239.
- VOSS, M., MOORE, J., HAUSER, M., GALLINAT, J., HEINZ, A., & HAGGARD, P. (2010). Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences. *Brain*, 133(10), 3104–3112.
- WARDAK, C. (2011). The role of the supplementary motor area in inhibitory control in monkeys and humans. *J Neurosci*, 31(14), 5181–5183.
- WEGNER, D. M. (2003). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- WEGNER, D. M., & WHEATLEY, T. (1999). Apparent mental causation. Sources of the experience of will. *Am Psychol*, 54(7), 480–492.
- WITTGENSTEIN, L. (1953). *Philosophical investigations*. Hoboken, NJ: Blackwell.

