



Action

Elisabeth Pacherie

► **To cite this version:**

Elisabeth Pacherie. Action. K. Frankish & W. Ramsey. The Cambridge handbook of cognitive science, Cambridge University Press, pp.92-111, 2012. <ijn_00778334>

HAL Id: ijn_00778334

https://jeannicod.ccsd.cnrs.fr/ijn_00778334

Submitted on 19 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

5 Action

Elisabeth Pacherie

5.1 Introduction

My arm rises. Is my arm rising something happening to me – say, a movement caused by a muscle spasm or by somebody pulling a string attached to my wrist – or is it my own doing – am I raising my arm? What does it mean to say that I am raising my arm intentionally? Must it be the case that a conscious intention to do so causes my arm to rise? How do I know that I am raising my arm, and does that knowledge differ from the knowledge you may acquire by observing me? The nature of action, action explanation, and agency are central issues in philosophical action theory and have been systematically explored in the last fifty years.

On the empirical side, with the emergence of cognitive neuroscience in the 1980s, motor cognition became a very active area of research. Work in the field of motor cognition aims at uncovering and understanding the mechanisms and processes involved in action specification and control. The efforts made to interpret anatomical and physiological evidence using cognitive theories and methods, including computational modeling, and, conversely, to test and refine cognitive models of normal motor cognition using functional neuroimaging and data from brain-damaged patients have resulted in a vast array of exciting discoveries and in provocative hypotheses about the cognitive structure of the processes and representations underpinning action.

The scientific study of action yields insights, distinctions, as well as descriptions of the causal mechanisms underlying action that go beyond what conceptual analysis, however sophisticated, could alone reveal. Results and ideas drawn from the scientific study of action can thus offer new sources of inspiration for philosophers, evidence which may help overcome longstanding difficulties or redraw the lines on the philosopher's map by challenging certain widely received assumptions. Conversely, careful philosophical analysis can also lead to more sober assessments of over-enthusiastic claims about what some recent empirical data show.

In recent years, the integration of philosophical with scientific theorizing has started to yield new insights. This chapter will survey some recent philosophical and empirical work on the nature and structure of action, on conscious agency, and on our knowledge of actions.

5.2 The nature of action and action explanation

One important debate that arose in the early sixties was concerned with whether the agent's reasons for his or her action were also the causes of the action. Following Wittgenstein, some philosophers (Anscombe 1963) argued that to explain why an agent acted as he or she did involved identifying the normative reasons that made the action intelligible in the agent's eyes and claimed that such normative explanations were different in kind from causal explanations. Others (Taylor 1964) similarly argued that explanations of actions are teleological explanations – in other words, explanations in terms of goals – and are as such not analyzable as causal explanations. In contrast, Davidson (1980, Ch. 1) argued that reason-explanations are causal explanations and did much to rebut the anti-causalist arguments that purported to show that reasons couldn't be causes. In particular, he pointed out that an agent may have several reasons to perform a certain action, but act only for one of those reasons. Challenging the non-causalists to provide an alternative explanation, he argued that what makes it true that the agent acts for this reason and not the other reasons he or she has is that this reason but not the others makes a causal contribution to the action. Similarly, most causalists will agree that reason-explanations for action are teleological but contend that teleological explanations are themselves kinds of causal explanations.

By reuniting the causal with the rational, the causalists opened the way for a naturalistic stance in action theory and thus for an integration of philosophical and scientific enquiries. The causal approach is today the dominant position in philosophical action theory. Broadly speaking, it considers that action is behavior that can be characterized in terms of a certain sort of psychological causal process. Yet, versions of the causal approach can take widely different forms depending on (1) what they take the elements of the action-relevant causal process to be, and (2) what part of the process they identify as the action. Thus, with respect to the first question, some theories countenance only beliefs and desires, while others view intentions, volitions or tryings as essential elements of the action-relevant causal structure. We can also distinguish three broad types of causal theories on the basis of their answer to the second question. On one view, one should characterize actions in terms of their causal power to bring about certain effects, typically bodily movements and their consequences. Accordingly, proponents of this view will tend to identify an action with mental events belonging to the earlier part of a causal sequence, such as tryings (Hornsby 1980). Conversely, one may hold that what distinguishes actions from other kinds of happenings is the nature of their causal antecedents. Actions will then be taken to be physical events (bodily movements and their consequences) with a distinctive mental cause. A third possibility is to consider actions as causal processes rather than just

causes or effects and to identify them with, if not the entire causal sequence, at least a large portion of it.

The earlier belief–desire versions of the causal theory, made popular most notably by Davidson (1980, Ch. 1) and Goldman (1970), held that what distinguishes an action from a mere happening is the nature of its causal antecedent, conceived as a complex of some of the agent’s beliefs and desires. One attraction of the belief–desire theory was its elegant simplicity. The theory took the belief–desire complex to both rationalize the action and cause it, thus simultaneously offering an account of the nature of actions – as events caused by belief–desire complexes – and an account of the explanation of intentional action as explanation in terms of the agent’s reasons for acting. Another important attraction of the theory was its ontological parsimony. It didn’t postulate any special type of mental events such as willings, volitions, acts of will, settings of oneself to act, tryings, etc. It did not even postulate intentions as distinct states, since on the theory, to say that somebody acted with a certain intention was just to say that his actions stood in the appropriate relations to his desires and beliefs.

However, it soon appeared that this simple version of the causal theory had serious shortcomings and remained incomplete in a number of important respects. First, as several philosophers have pointed out, including Davidson himself (Davidson 1980, Ch. 5; Bratman 1987), the relational analysis of intentions is inapplicable to intentions concerning the future, intentions which we may now have, but which are not yet acted upon, and, indeed, may never be acted upon. Acknowledging the existence of future-directed intentions forces one to admit that intentions can be states separate from the intended actions or from the reasons that prompted the action. But, as Davidson himself notes, once this is admitted, there seems to be no reason not to allow that intentions of the same kind are also present in all or at least most cases of intentional actions.

Second, it was also pointed out (Brand 1984; Searle 1983) that the belief–desire theory does not account for “minimal” actions, i.e., actions that are performed routinely, automatically, impulsively or unthinkingly. To borrow an example from Searle (1983), suppose I am sitting in a chair reflecting on a philosophical problem, and I suddenly get up and start pacing about the room. Although my getting up and pacing about are actions of mine, no antecedent belief–desire complex prompted me to do so. The act was unpremeditated and spontaneous. Thus, it may be doubted whether being caused by a belief–desire complex is a necessary condition for an event to qualify as an action.

What these two objections suggest is that actions come in various grades, from routinely performed low-level purposive behavior to deliberately undertaken and consciously preplanned actions, and thus that their psychological structure may be more or less rich.

A third objection to the belief–desire version of the causal theory is that it doesn’t have the resources to exclude aberrant manners of causation. This is the notorious problem of causal deviance or waywardness. Here’s an example from Mele:

Ann wants to awaken her husband and she believes that she may do so by making a loud noise. Motivated (causally) by this desire and belief, Ann may search in the dark for a suitable noise-maker. In her search, she may accidentally knock over a lamp, producing a loud crash. By so doing, she may awaken her husband, but her awakening him in this way is not an intentional action. (Mele 2002, pp. 21–2)

As this example illustrates, not every causal relation between seemingly appropriate mental antecedents and resultant events qualifies the latter as intentional actions. The challenge then is to specify the kind of causal connection that must hold between the antecedent mental event and the resultant behavior for the latter to qualify as an intentional action.

A fourth, related problem, concerns the explanation of failed actions. Some actions fail because some of the agent’s beliefs are false. Thus, John may fail to turn the light on because he was wrong to believe that the switch he pressed commanded the light. The causal theory can account for failures of this kind, for it claims that the (non-accidental) success of an action depends on the truth of the beliefs figuring in the motivating belief–desire complex. Yet, as Israel, Perry, and Tutiya (1993) point out, the failure of an action cannot always be traced back to the falsity of a motivating belief. Here’s their example. Suppose Brutus intends to kill Caesar by stabbing him. His beliefs that Caesar is to his left and that stabbing Caesar in the chest would kill him are both true, and yet Brutus fails to kill Caesar because he makes the wrong movement and misses Caesar completely. This is what they call the “problem of the wrong movement”: when the agent’s beliefs are correct, what ultimately accounts for the success or failure of an intended action are the bodily movements performed. If we consider that a theory of action explanation should aim at explaining the actual action, not just the attempt or volition, we should be ready to include in the motivating complex cognitions pertaining to movements. The motivating complex as it is conceived in the standard account is thus fundamentally incomplete, leaving a gap between the motivating cognitions and the act itself.

The various revisions and refinements the causal theory of action has undergone in the last three decades can be seen as attempts to overcome some of these difficulties and shortcomings. In particular, many philosophers have found it necessary to introduce a conception of intentions as distinctive, *sui generis*, mental states. They argue that intentions have their own complex and distinctive functional role and form an irreducible kind of psychological state, on a par with beliefs and desires. Thus, Bratman (1987) stresses

three functions of intentions. First, they are *terminators of practical reasoning* in the sense that once we have formed an intention to *A*, we will not normally continue to deliberate whether to *A* or not; in the absence of relevant new information, the intention will resist reconsideration. Second, intentions are also *prompters of practical reasoning*, where practical reasoning is about means of *A*-ing. This function of intentions thus involves devising specific plans for *A*-ing. Third, intentions also have a *coordinative function* and serve to coordinate the activities of the agent over time and to coordinate them with the activities of other agents.

Philosophers also typically point out further functions of intentions (Brand 1984; Mele 1992). Intentions are also responsible for triggering or initiating the intended action (*initiating function*) and for guiding its course until completion. An intention to *A* incorporates a plan for *A*-ing, a representation or set of representations specifying the goal of the action and how it is to be arrived at. It is this component of the intention that is relevant to its *guiding function*. Finally, intentions have also been assigned a *control function*, involving a capacity to monitor progress toward the goal and to detect and correct deviations from the course of action as laid out in the guiding representation.

The first three functions of intentions just described – their roles as terminators of practical reasoning about ends, as prompters of practical reasoning about means and as coordinators – are typically played by intentions in the period between their initial formation and the initiation of the action. By contrast, the last three functions (initiating, guiding, and controlling) are played in the period between the initiation of the action and its completion. Attention to these differences has led a number of philosophers to develop dual-intention theories of action. For instance, Searle (1983) distinguishes between prior intentions and intentions-in-action, Bratman (1987) between future-directed and present-directed intentions, and Mele (1992) between distal and proximal intentions. In all cases, an intention of the former type will only eventuate into action by first yielding an intention of the latter type.

Dual-intention theories make available new strategies for dealing with the difficulties listed earlier. To begin with, they open up new prospects toward a solution to the problem of minimal actions (that many actions do not seem to be preceded by any intention to perform them). According to dual-intention theories, all actions have proximal intentions, but they need not always be preceded by distal intentions (from now on, I use Mele's terminology). For instance, when, reflecting on a philosophical problem, I start pacing about the room, I do not first engage in a deliberative process that concludes with a distal intention to pace; rather my pacing is initiated and guided by a proximal intention formed on the spot. Automatic, spontaneous or impulsive actions may then be said to be those actions that are caused by proximal intentions but are not planned ahead of time.

Dual-intention theories also provide at least a partial answer to the problem of causal deviance. They suggest that for intentions to cause actions in the right way for them to count as intentional, two constraints should be met. First, the intended effect must be brought about in the way specified by the plan component of the intention. Second, it must also be the case that the causal chain linking the distal intention to the resultant bodily behavior include relevant proximal intentions.

5.3 Motor cognition

Although dual-intention theories sound more promising than the earlier belief-desire theory, more needs to be said about the ways intentions carry out their functions and about the nature of their contents. First, if proximal intentions are to be regarded as playing an essential role in the initiation of all cases of action, one should identify the features of proximal intentions that allow them to play this role. Second, we need an account of the guidance and monitoring functions of proximal intentions. Third, in cases where the agent is acting on his distal intention, there must be an appropriate transition between the distal intention and the proximal intention, and we need to clarify what constitutes an appropriate transition.

Work in the field of motor cognition is highly relevant to these issues. This field integrates research techniques and methods from cognitive psychology, behavioral neuroscience, and computational modeling in an attempt to provide a unified approach to the central questions of the organization of action, the nature and role of the different representations involved in the generation of action, and the contributions of different brain structures to the planning and execution of movement. Here, I will concentrate on the functional architecture of motor cognition, introducing some of the theoretical concepts, models, and hypotheses that play a central role in current thinking in the motor domain and are of particular relevance for philosophical theorizing on action.

Work on motor physiology started at the end of the nineteenth century and was long dominated by the sensory-motor theory of action generation that conceived of actions as reactions to changes in the external environment and as essentially a matter of movements and the muscles that power them. Thus Sherrington, the famous British neurophysiologist, considered the reflex action as the elementary unit of behavior and thought that all coordinated action was constructed through a process of sequential combination, where reflexes were chained into behavioral sequences in such a way that feedback from one movement stimulated the next in the sequence (Sherrington 1947). This view of complex actions as associative chains left little role for cognitive processes in the organization of action.

In the early 1950s Karl Lashley (1951) launched an attack against this view and argued that the action sequence is guided by plans and motor programs.

He pointed out that complex action sequences are characteristic of human behavior and that humans are remarkably adept at learning new skills and rearranging elementary movements to produce new action sequences. The fact that the same elementary movements can occur in different orders raises an obvious problem for the idea of serial chaining since a given movement may be followed by different movements on different occasions. Another argument in favor of the central organization of action (as opposed to peripheral chaining) comes from the fact that we do not simply react to external events but also actively initiate interactions with our environment.

Centralism, the idea that voluntary actions are largely driven by central internal representations rather than by external events is one of the central tenets of contemporary theories of action generation. As Jeannerod (1997, 2006) points out, to be capable of internally generated purposive action, an organism must have internal models of how the external world is, how it will be modified by the action of the organism, and how the organism itself will be modified by this action. The modern idea of internal models had several precursors. One of them is the idea of a homeostatic device, where the signals that initiate a process originate from a discrepancy between a central signal and an input signal, the former corresponding to a fixed inbuilt reference value for some parameter, the other to the current value of the parameter. Homeostatic systems draw attention to the role of endogenous factors and imply the existence of a certain form of representation or stored knowledge of the reference value of a parameter. Another precursor of internal models is the concept of efference copy proposed by Von Holst and Mittelstaedt (1950). The idea is that when the motor centers send a motor command to the peripheral nervous system to produce a movement, they also send a copy of this command to other centers that can in this way anticipate the effect of the motor command. (A motor signal from the central nervous system to the periphery is called an *efference*, and a copy of this signal is called an *efference copy*.) The notion of an efference copy is of particular interest for two reasons. First, it is a centrally generated signal, and this suggests that the central nervous system can inform itself directly about its current state and activity without a detour through peripheral reafferences. Second, it constitutes an elementary instance of expectation or anticipation, where an internal model of forthcoming sensory experience arises in advance of actual feedback.

The concept of internal models was further developed by engineers who proposed computational theories incorporating the idea of control strategies based on internal models and have applied this approach in the fields of robotics, neural networks, and adaptive control. There is now growing evidence that similar strategies are used in human motor control (e.g., Jeannerod 1997; Frith, Blakemore, and Wolpert 2000).

Current computational theories of human motor control appeal to two main kinds of internal models, forward and inverse models, as illustrated in

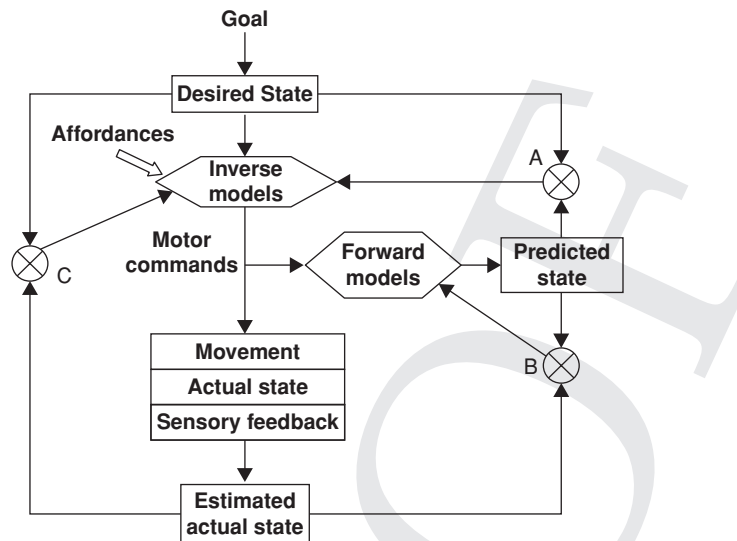


Figure 5.1 The basic components of a motor control system based on internal models. Adapted from Frith *et al.* 2000.

Figure 5.1. In a nutshell, an inverse model (or controller) computes the commands for achieving a desired state given the current state of the system and of the environment. An efference copy of these commands is fed to a forward model (also called predictive model) that represents the causal flow of a process in a system and can thus generate a prediction of the consequences of performing these commands. Of special interest is the idea that the control of action depends in a large part on the coupling of inverse and forward models through a series of comparators, i.e., mechanisms that compare two signals and use the result of the comparison for various kinds of regulation.

A first kind of comparator (labelled A in Figure 5.1) takes as input representations of the desired state and of the predicted state and sends an error signal to the inverse model if a difference is found. Such a mechanism can be used to maintain accurate performance in the presence of feedback delays. It can also be used for mental practise and planning, as forward models can predict the sensory outcome of an action without the action being actually carried out. A second kind of comparator mechanism (labelled B in Figure 5.1) compares the predicted consequences of a motor command with its actual consequences. The result of this comparison can be used to update the forward model and improve its functioning. It can also be used to filter sensory information and to distinguish the component that is due to self-movement from that due to changes in the world (Blakemore, Wolpert, and Frith 1999). Finally, a third kind of comparison is between desired state and actual feedback (labelled C in Figure 5.1). Errors derived from the difference between the

desired state and the actual state can be used to update the inverse models and improve performance. This kind of comparison is therefore important for motor learning.

A third key tenet of current theorizing on motor cognition, besides the idea of central representations and the idea of control structures involving internal and external feedback loops, is the idea that action is hierarchically organized. The organization of action is commonly thought of as a functional hierarchy comprising three main levels, corresponding to the progressive specification of the action to be performed. At the highest level, action representations represent the whole action as a unit, in terms of its overarching goal and of the sequence of steps or subgoals needed to achieve that goal. At this level, the action is represented in a rather abstract, typically conceptual, format. The second level is concerned with the implementation of each step in the action plan and involves selecting an appropriate motor program given the immediate goal and contextual information about the current state of the agent and the current state of its environment. In other words, processes at this level are in charge of anchoring the successive steps of the action plan in the current situation and of selecting appropriate motor programs. Finally, once a motor program has been selected, the exact values of its parameters must still be set. This is done at the third level, where incoming sensory information about external constraints is used to specify these values.

This distinction of three levels is an oversimplification and should be qualified in several ways. First, the organization within each level can itself be decomposed into hierarchical stages. Second, the distinction between the first and second level is not always sharp. A given action may be planned to a greater or a lesser extent. Typically, how much is planned at the highest level depends on the agent's expertise. For instance, while the novice tennis player intent on performing a topspin serve may have to represent all the steps involved in performing such a serve in advance of acting, the expert tennis player need only represent his action as a topspin serve at the planning level, having already built through intensive training an appropriate motor program where these steps are stored. Third, talk of a hierarchical organization and of a series of levels may give the impression that the processing steps must be ordered serially – that planning must be over before programming starts, and that programming in turn must be over before the execution starts. As Jeannerod (1997) points out, however, activation in the cortical areas thought to correspond to the various levels of organization occurs simultaneously and the existence of a sequence can only be detected statistically. Yet, the existence of parallel processing in the motor system does not contradict the idea of hierarchy of levels. A hierarchy between levels implies degrees of specialization for each level but it does not imply a sequential order of activation.

One important source of evidence for the hierarchical organization of actions comes from neuropsychology, where lesions in different brain areas

may lead to different types of impairments of motor cognition. Thus, patients with optic ataxia produce inaccurate reaching movements toward a target or object in space as well as inaccurate grasping of objects with incorrect orientating of the hand and inadequate pre-shaping with respect to the shape and size of the object (Rossetti, Vighetto, and Pisella 2003). Their visuomotor impairment affects the bottom level of the organization of action, concerned with appropriately setting the parameters of the selected motor programs. In contrast, patients suffering from ideomotor apraxia (Heilman and Rothi 1993) have no problem reaching for and grasping objects and can describe what their functions are, but they are not able to manipulate them according to their function. Their deficit relates to the second level of action organization: they seem to have lost the motor programs associated with various kinds of objects. Finally, patients with utilization behavior reach out and automatically use objects in an instrumentally correct manner that is inappropriate for the particular context (Lhermitte 1983). For instance, a patient seeing a pair of glasses placed in front of him may pick it up and put it on. Moreover, if a second and then a third pair of glasses are placed in front of him, he will put them on and will end up wearing all three. In contrast, when they lack external stimulation to steer them into action, these patients exhibit mental inertia and apathy. They seem to be impaired at the highest level of action organization: they have lost the capacity to generate and act on endogenous intentions and, as a result, to inhibit stimulus-driven actions that are normally kept in check by endogenous plans.

The three notions of central representations of action, control structures making use of internal models, and hierarchical organization of action are highly relevant to the concerns of philosophers of action. Firstly, the idea of a hierarchical organization of action representations and control structures helps flesh out the idea that actions come in various grades, from minimal, automatic, highly routinized actions to carefully preplanned actions with long-term and complex goals, and can have a psychological structure whose richness varies accordingly. The two highest levels in this hierarchy echo the distinction of distal and proximal intentions proposed by dual-intention theorists. However, with but a few exceptions (Pacherie 2008) philosophers ignore the third and lowest level of the hierarchy. Secondly, the idea that action representations are associated with control structures involving inverse and forward models coupled through comparators helps make sense of the idea that representations of actions can be both teleological and causal. They are representations of action goals that both cause action specification and execution and control progress toward the goal through internal and external feedback loops. Thirdly, careful attention to the way action representations control the performance of the agent may also give us a solution to problems of causal deviance. Yet, as we will now see, recent empirical work can also yield results that appear to challenge deeply entrenched philosophical assumptions.

5.4 Conscious agency

Libet (1985) suggested that the results of his studies on brain activity during the preparation of voluntary acts seriously questioned the idea that conscious intentions have any causal role in the initiation of action and therefore threatened the notion of free-will as traditionally understood. More recently, Wegner's psychological experiments led him to claim that the conscious will is an illusion (Wegner 2002). These attacks on the traditional view of the role of conscious agency did much to reawaken the interest of philosophers in the phenomenology of action. At the same time, further empirical investigations aimed at probing in more detail the phenomenology of action and its disorders have started yielding a wealth of new data, suggesting that extreme skepticism vis-à-vis conscious agency may rest in part on too simplistic a view of the phenomenology of agency.

In his famous studies, Libet (1985) asked subjects to move a hand at will and to note when they felt the urge to move by observing the position of a dot on a special clock. While the participants were doing this, the experimenters recorded their readiness potential, i.e., the brain activity linked to the preparation of movement. What they found was that the onset of the readiness potential predated the conscious awareness of the urge to move by about 350 milliseconds, while the actual onset of movement measured in the muscles of the forearm occurred around 150 milliseconds after conscious awareness. Libet and others have claimed that these results provide evidence in favor of a skeptical attitude toward conscious mental causation: since the conscious awareness of the urge to move occurs much later than the onset of the brain activity linked to the preparation of movement, it could play no causal role in the production of the intentional arm movement. Libet himself suggested that consciousness may still intervene and veto the unconsciously initiated action, providing a kind of conscious "free won't."

Several philosophers have criticized Libet's interpretation of the bearing of his experiments on conscious agency and free will. First, it is worth noting that although the conscious urge to move may lag behind the onset of brain activity, it still precedes the actual onset of movement. Libet's interpretation of his finding is premised on the view that only the initial element in a causal chain, i.e., only a cause uncaused, may genuinely qualify as a cause. Yet, the notion of a cause uncaused is metaphysically dubious and certainly hard to square with a naturalistic stance. A conscious mental state may play a causal role in the production of an action even though it doesn't trigger the whole causal process. If it makes a difference whether or not a causal chain contains conscious mental states as elements, and in particular if there are differences in the kinds of actions that can be the outcome of such chains or in the conditions in which such actions can be successfully performed, then it is fair to say that conscious mental states make a difference and are

causally efficacious. One may also note that the unconscious processes that precede conscious awareness are not themselves uncaused and that, by parity of reasoning, Libet should also deny that they initiate the action.

Second, as Mele (2003) points out, it is unclear whether the readiness potential constitutes the neural substrate of intentions or decisions rather than of desires or urges. If the latter, no one should be surprised to find that desires precede conscious intentions, and finding that we have such desires does not commit us to acting upon them. For all Libet has shown, it may be that another conscious act is necessary before the event associated with the readiness potential leads to action. Third, Libet's analysis focuses on proximal intentions (the proximal causes of overt behavior, whose content in this case may be expressed as "I flex my wrist thus and thus now"), but it neglects distal intentions (whose content may be expressed as "I will flex my wrist when I feel the urge"). Yet, it is quite implausible that the participants in his studies would have produced hand movements at will unless they had formed the distal intention to do so in compliance with the experimenter's instructions. This suggests that distal intentions are not causally inert.

Wegner's claim that the conscious will is an illusion would seem, if empirically warranted, even more damaging to our traditional concept of will and conscious agency than Libet's findings. One line of argument Wegner advances in favor of this claim appeals to dissociations, i.e., cases in which agency and the experience of agency come apart. For instance, in his I-spy experiment (Wegner and Wheatley 1999), a participant and a confederate of the experimenter have joint control of a computer mouse that can be moved over any one of a number of pictures on a screen. When participants had been primed with the name of an item on which the mouse landed, they showed an increased tendency to self-attribute the action of stopping on that object (when in fact the stop had been forced by the confederate). In other words, they experienced conscious will for an action they had not actually controlled. Wegner also argues that many apparently occult phenomena, such as table turning and the ouija board, are instances of the reverse dissociation: the agents in question are doing things that they are not aware they are doing. Wegner seems to think that since the mechanisms responsible for the phenomenology of agency are fallible, we have no reason to think that our experience of agency can ever be trusted. This inference appears less than compelling. To show that the experience of willing is not always errorless is certainly not to show that it is always in error. Indeed, it may well be highly reliable most of the time.

Two further lines of argument for the illusory character of conscious will comes from Wegner's account of how the experience of conscious will is generated, what he calls the theory of apparent mental causation. According to this theory, conscious will is experienced when we infer, correctly or not, that our thought has caused our action. We draw such an inference when

we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. In actual fact, however, our actions spring from subpersonal causal processes and the conscious ideas that we mistakenly experience as their causes are themselves caused by subpersonal processes which may have only indirect links to the subpersonal processes causing the action.

Wegner's thought here seems to be that the real causal work is done by subpersonal processes and that subpersonal explanations pre-empt personal-level explanations. However, as Bayne (2006) points out, an alternative to this eliminativist position is to see these explanations as complementary. One might regard subpersonal explanations as explaining how intentional agency is realized rather than explaining it away. Wegner also seems to think that the conscious will is an illusion insofar as our experience is inferentially mediated rather than being a direct report of the processes whereby action is produced. If "direct report" is taken to mean that no subpersonal processes or inferential mechanisms of any kind are involved in generating the experience of agency, it is far from clear that a direct report view is a plausible view of the experience of agency or of any other kind of conscious experience. More importantly, Wegner offers no good reason for thinking that the experience of agency could be reliable only if it were a direct readout of action-production processes.

As Jeannerod and others have demonstrated, our conscious access to the representations and processes involved in action specification and control gets more and more limited as we go down the hierarchy of action organization, with the processes and representations at the lowest level being typically unavailable to consciousness. Thus, Wegner may well be right that the experience of conscious will is typically not a direct phenomenal readout of action-production processes and must be theoretically mediated. Yet, there are reasons to doubt that, as Wegner's model suggests, the experience of conscious will arises solely or primarily when there is a match between a prior thought and an observed action. First, prior thoughts or awareness thereof do not seem to be necessary for the sense of agency. On many occasions, we cannot remember what our prior intentions were and yet do not disown our actions. Furthermore, many of our actions, impulsive, routine or automatic, are not preceded by conscious previews and yet we own them. Second, awareness of a match between a prior thought and an action does not seem sufficient for a sense of agency. For instance, schizophrenic patients suffering from delusions of control may lack a sense of agency for a given action despite being aware that what they are doing matches their prior intention (Frith *et al.* 2000).

Recent empirical work suggests that other types of matches than just the match between a prior intention and an observed action play a role. One such match is between a voluntary movement and its consequences. Haggard and colleagues (Haggard and Clark 2003; Moore and Haggard 2008) have shown that when a voluntary act (a button press) causes an effect (a tone), the

perceived time of initiating the act is closer to the perceived time of the effect. Specifically, the action (the button press) is shifted forward in time toward the effect it produces, while the effect is shifted backward in time toward the action that produces it. Haggard calls this phenomenon intentional binding.

Several lines of evidence suggest that intentional binding probably derives from predictive mechanisms of action control and is based on the comparison between the predicted sensory consequences of a voluntary movement and its actual sensory consequences. First, intentional binding depends critically on the presence of voluntary movement and requires an efferent signal. When similar movements and auditory effects occur involuntarily or when transcranial magnetic stimulation (TMS) is used to insert occasional involuntary movements of the right finger at a time when the subject intends to press the button but has not yet done so, the binding effect is reversed and cause and effect are perceived as further apart in time than they actually are. Second, intentional binding requires reliable relations between actions and effects and largely depends on the degree of discrepancy between the predicted and actual sensory feedback (Moore and Haggard 2008).

Haggard suggests that the same neural mechanism that produces intentional binding of actions also produces the sense of agency we experience for our actions and that, therefore, intentional binding may be an implicit measure of the sense of agency. Indeed, studies by Sato and Yasuda (2005) show that the same factors that modulate intentional binding also modulate the sense of self-agency subjects experience for the action.

Like Wegner, Haggard proposes a matching model of the experience of agency. His findings suggest, however, that the processes through which the sense of agency is generated are much more closely linked to the processes involved in the specification and control of action than Wegner thinks. He takes the experience of agency to depend primarily on the degree of match between the sensory consequences of an action as predicted by the motor system and its actual sensory consequences rather than on a match between a prior conscious thought and an action. Yet, neither a top-down inferential approach *à la* Wegner nor a purely bottom-up approach involving only subpersonal processes is entirely satisfactory if taken in isolation. Rather than choosing between them, several authors in the field now argue for theoretical integration and a multiple-aspects approach to the problem (Bayne and Pacherie 2007; Gallagher 2007; Pacherie 2008; Synofzik, Vosgerau, and Newen 2008).

5.5 Knowledge of actions and intentions

It is commonly held that whereas our knowledge of the intentions and actions of others involves inferring their mental states from their observed behavior,

we have direct knowledge of our own actions and intentions without having to rely on observation and inference. This supposed asymmetry gives rise to a skeptical worry concerning the very possibility of knowledge of others' intentions and actions. If the process through which we make mental attributions to others is one of theoretical inference, where we observe their behavior and infer the mental state thought to be its causal antecedent, then it seems in principle possible that the theory upon which the inferences are based is incorrect and therefore that any given attribution of a mental state to others could be false.

This way of conceiving of the problem of other minds is a consequence of a Cartesian picture of the mind and its relation to bodily behavior. According to this picture, what confer intentional properties to behavior are its inner mental accompaniments and causes. In other words, nothing intrinsic distinguishes a mere bodily happening from a piece of intentional behavior; the difference is one of causal antecedents. Since internal mental causes can't be directly observed, they must be inferred, thus leaving open the possibility that the inference be incorrect. In this respect, versions of the causal theory of actions that take actions to be bodily movements with a distinctive mental cause are still very much in the grip of the Cartesian picture.

Alternatively, it can be argued that behavior and mentality are much more integrated than the Cartesian picture suggests and that the actions and intentions of others can be, at least to some extent, available to experience in their own right, rather than having to be inferred on the basis of behavioral proxies. This alternative view rests on three complementary claims: (1) that intentional bodily behavior has distinctive intrinsic characteristics, (2) that we are perceptually sensitive to these characteristics, and (3) that the internal representations we form when observing intentional behavior are similar to those we form when performing intentional behavior.

A large body of empirical evidence now exists in favor of these three claims. Intentional behavior has been shown to have distinctive observable properties, a distinctive kinematics, and a dynamics that bears systematic relations to features of the situation in a way that non-intentional behavior does not. There is also ample empirical evidence that we are perceptually attuned to these unique characteristics of intentional behavior. Perceptual sensitivity to human motion is already present in infants aged between 3 and 5 months (Bertenthal, Proffitt, and Cutting 1984) and seems therefore to be innate or to develop very early. Habituation studies also indicate that infants are sensitive to the goal-directed structure of certain actions by the time they are 5–6 months of age (Woodward 2005). There is also extensive evidence that adult subjects can quickly and reliably recognize movement patterns of walking, cycling, climbing, dancing, etc., from kinematic information alone (Johansson 1973).

Recent neurological studies have yielded a set of important results on mirroring processes. In a series of single-neuron recording experiments on macaque monkeys designed to investigate the functional properties of neurons in area F5, Rizzolatti and his colleagues discovered so-called mirror neurons, i.e., sensorimotor neurons that fire both during the execution of purposeful, goal-related actions by the monkey and when the monkey observes similar actions performed by another agent (Fogassi and Gallese 2002; Rizzolatti and Craighero 2004). In addition, a large body of neuroimaging experiments have investigated the neural networks engaged during action generation and during action observation in humans, revealing the existence of an important overlap in the cerebral areas activated in these two conditions (for reviews, see Grèzes and Decety 2001; Jeannerod 2006). These results have been interpreted as support for the existence of a process of motor simulation or motor resonance whereby the observation of an action activates in the observer an internal representation of the action that matches the representation of the action activated in the brain of the performer. By linking self and other through a unique framework of shared representations of action, mirror systems would allow one to directly understand the actions of others. The nature and extent of the understanding of others that mirroring processes can provide has given rise to an intense debate, with some theorists seeing them as the fundamental neural basis of human social cognition (e.g., Gallese 2007), while others hold more deflationary views (e.g., Jacob 2008).

5.6 Conclusion

In the last decades, philosophers have developed sophisticated conceptual frameworks for thinking about the psychological structures of action. During the same period, empirical investigations have led to a better understanding of motor cognition. Integrating these complementary insights yields the prospects of a more comprehensive picture of action from deliberation and planning down to motor execution. This integrative approach still needs to proceed further. Philosophers haven't yet fully assessed the implications of empirical findings on action preparation and control processes for their views of the nature of intentional action. Conversely, neuroscientists have only recently started investigating how and where the brain stores distal intentions (Haynes *et al.* 2007). Recent controversies on free will and conscious agency also suggest that progress on these issues may depend on further collaborative efforts by philosophers and scientists. Finally, it remains to be seen how much of social cognition has its neural bases in mirroring processes. Here, one promising new area of investigation is joint action and the cognitive and neural processes that support it (Sebanz, Bekkering, and Knoblich 2006).

Further reading

- Jeannerod, M. (2006). *Motor Cognition*. Oxford University Press. A stimulating and up-to-date synthesis of work on motor cognition encompassing neuropsychology, neurophysiology, philosophy, neuroimaging, comparative neurobiology, and clinical studies by a foremost researcher in this new interdisciplinary field.
- Mele, A. R. (ed.) (1997). *The Philosophy of Action*. Oxford University Press. A selection of some of the most influential essays on the major contemporary issues in the philosophy of action. The introductory essay by A. Mele provides a clear guide to the current debates.
- Moya, C. J. (1990). *The Philosophy of Action: An Introduction*. Cambridge: Polity Press. A clear and concise introduction to the philosophy of action, accessible to readers without special philosophical training.
- Pockett, S., Banks, W. P., and Gallagher, S. (eds.) (2006). *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*. Cambridge, MA: MIT Press. This multidisciplinary collection continues the debate over whether consciousness causes behavior or plays no functional role in it, approaching the question from both empirical and theoretical perspectives. Contributors also examine the effect recent psychological and neuroscientific research could have on legal, social, and moral judgments of responsibility and blame.
- Rizzolatti, G. and Sinigaglia, C. (2007). *Mirrors in the Brain: How Our Minds Share Actions, Emotions, and Experience*. Oxford University Press. Jointly written by one of the discoverers of mirror neurons and a philosopher, this very readable book provides a systematic overview of mirror neurons and investigates the role of mirroring processes in action understanding, imitation, language, and the sharing of emotions.
- Roessler J. and Eilan, N. (eds.) (2003). *Agency and Self-Awareness*. Oxford University Press. A collection of essays by philosophers, psychologists, neuropsychologists, and neuroscientists on consciousness of action, its role in the control of intentional action, and its contribution to self-awareness.
- Sebanz, N. and W. Prinz (eds.) (2006). *Disorders of Volition*. Cambridge, MA: MIT Press. In this collection of essays, philosophers, psychologists, neuroscientists, and psychiatrists seek to advance our understanding of the processes supporting voluntary action by addressing the will and its pathologies from both theoretical and empirical perspectives.

References

- Anscombe, G. E. M. (1963). *Intention* (2nd edn.). Ithaca, NY: Cornell University Press.
- Bach, K. (1978). A representational theory of action, *Philosophical Studies* 34: 361–79.
- Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will, in S. Pockett, W. P. Banks, and S. Gallagher (eds.), *Does*

- Consciousness Cause Behavior? An Investigation of the Nature of Volition* (pp. 169–85). Cambridge, MA: MIT Press.
- Bayne, T. and Pacherie, E. (2007). Narrators and comparators: The architecture of agentive self-awareness, *Synthese* 159: 475–91.
- Bertenthal, B. I., Proffitt, D. R., and Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions, *Journal of Experimental Child Psychology* 37: 213–30.
- Blakemore, S.-J. Wolpert, D. M., and Frith, C. D. (1999). Spatiotemporal prediction modulates the perception of self-produced stimuli, *Journal of Cognitive Neuroscience* 11: 551–9.
- Brand, M. (1984). *Intending and Acting*. Cambridge, MA: MIT Press.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Davidson D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.
- Fogassi, L. and Gallese, V. (2002). The neural correlates of action understanding in non-human primates, in M. I. Stamenov and V. Gallese (eds.), *Mirror Neurons and the Evolution of Brain and Language* (pp. 13–35). Amsterdam: John Benjamins.
- Frith, C. D., Blakemore, S.-J., and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action, *Philosophical Transactions of the Royal Society of London B* 355: 1771–88.
- Gallagher, S. (2007). The natural philosophy of agency, *Philosophy Compass* 2: 347–57.
- Gallese, V. (2007). Before and below “theory of mind”: Embodied simulation and the neural correlates of social cognition, *Philosophical Transactions of the Royal Society of London B* 362: 659–69.
- Goldman, A. (1970). *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.
- Grèzes, J. and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of actions: a meta-analysis, *Human Brain Mapping* 12: 1–19.
- Haggard, P. and Clark, S. (2003). Intentional action: Conscious experience and neural prediction, *Consciousness and Cognition* 12: 695–707.
- Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness, *Nature Neuroscience* 5: 382–5.
- Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R. E. (2007). Reading hidden intentions in the human brain, *Current Biology* 17, 323–8.
- Heilman K. M. and Rothi L. J. G. (1993). Apraxia, in K. M. Heilman and E. Valenstein (eds.) *Clinical Neuropsychology* (pp. 141–163, 3rd edn). Oxford University Press.
- Hornsby, J. (1980). *Actions*. London: Routledge and Kegan Paul.
- Israel, D., Perry, J., and Tutiya, S. (1993). Executions, motivations and accomplishments, *Philosophical Review* 102: 515–40.

- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language* 23: 190–223.
- Jeannerod, M. (1997). *The Cognitive Neuroscience of Action*. Oxford: Blackwell.
- (2006). *Motor Cognition*. Oxford University Press.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis, *Perception and Psychophysics* 14: 201–11.
- Lashley, K. S. (1951). The problem of serial order in behavior, in L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior* (pp. 112–36). New-York: Wiley.
- Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes, *Brain* 106: 237–55.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action, *Behavioral and Brain Sciences* 8: 529–66.
- Mele A. R. (1992). *Springs of Action*. Oxford University Press.
- (2002). Action, in L. Nadel (ed.), *Encyclopedia of Cognitive Science* (pp. 20–23). London: MacMillan.
- (2003). *Motivation and Agency*. Oxford University Press.
- Moore, J. and Haggard, P. (2008). Awareness of action: Inference and prediction, *Consciousness and Cognition* 17: 136–44.
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework, *Cognition* 107: 179–217.
- Perenin, M. T. and Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. I: Different aspects of the deficit in reaching for objects, *Brain* 111: 643–74.
- Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system, *Annual Review of Neuroscience* 27: 169–92.
- Rossetti, Y., Vighetto, A., and Pisella, L. (2003). Optic ataxia revisited: Immediate motor control versus visually guided action, *Experimental Brain Research* 153: 171–9.
- Sato, A. and Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of ownership, *Cognition* 94: 241–55.
- Searle, J. R. 1983. *Intentionality*. Cambridge University Press.
- Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: Bodies and minds moving together, *Trends in Cognitive Science* 10: 70–6.
- Sherrington, C. S. (1947). *The Integrative Action of the Nervous System*. New Haven: Yale University Press.
- Synofzik, M., Vosgerau, G., and Newen, A. (2008). Beyond the comparator model: A multifactorial two step account of agency, *Consciousness and Cognition* 17: 411–24.
- Taylor, C. (1964). *The Explanation of Behaviour*. London: Routledge and Kegan Paul.
- Von Holst, E. and Mittelstaedt, H. (1950). Das Reafferenz Prinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie, *Naturwissenschaften* 37: 464–76.

- Wegner, D. M. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wegner, D. M. and Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will, *American Psychology* 54: 480–92.
- Woodward, A. L. (2005). Infants' understanding of the actions involved in joint attention, in N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds.), *Joint Attention, Communication, and Other Minds* (pp. 85–109). Oxford University Press.

PROOF