

5 Neuroscience and Agent-Control

Philip Pettit

Neuroscientific and related studies have sharpened the debate about how we, creatures composed out of neurally organized matter, can exercise the sort of agent-control over action that makes it possible for us to be held responsible for them. In this chapter I look briefly at some results that sharpen this problem, argue that the problem arises because we think of agent-control on what I call the act-of-will picture, and then try to show that by moving to an alternative conception of agent-control we can resolve the difficulty; this conception of agent-control I have already defended elsewhere (Pettit and Smith 1996; Pettit 2001a, b). The scientific results suggest not that agent-control is an illusion but that having agent-control is best conceived in the manner I favor, or in some closely related fashion.

The chapter is in five sections. In the first three, I describe and present the challenge, identify an assumption that it presupposes and propose an alternative, and then argue that under this alternative assumption the challenge can be satisfactorily met. In the fourth section, I look more generally at the way the alternative picture relates to the standard one it displaces. And in the fifth, I offer a comment on the sort of lesson that neuroscience teaches us in this and other cases.

The Challenge

One of the standard assumptions in traditional philosophical lore is that actions are subject to agent-control if and only if they are the products of acts of will. Acts of will, on this approach, satisfy three crucial conditions. First, they reveal the mind of the agent in a pure or direct manner, not in the manner of the actions they prompt; they cannot be actions themselves, since they would then require prior acts of will and a regress would open up. Second, they are essentially capable of being brought to awareness by

the agent, even in the case of habitual behavior; they are so intimately bound to the agent that no ignorance of their reality is inevitable and no error about their identity possible. And third, they are acts that the agent can initiate on demand, as choice is called for; the agent can will or not will, just as fancy goes.

The act-of-will picture is highly plausible from our point of view as agents. We know what it is just to will something; we know that we are usually quite capable of telling what our will is, and what it is not; and we feel, perhaps as deeply as we feel anything, that we can display our capacity in these regards by consciously making a choice on demand: say, raising the right hand rather than the left, or the left rather than the right. 'I am in charge, so the intuition has it, and being in charge, being in agent-control, means that for the relevant domain—however restricted this may be—things happen according to my will. My will is where I am, and what I do, consciously or otherwise, I do by willing it.'

The picture is beautifully caught in Iain McEwan's novel, *Atonement*, when the twelve-year-old Briony reflects as follows:

She raised one hand and flexed its fingers and wondered, as she had sometimes before, how this thing, this machine for gripping, this fleshy spider on the end of her arm, came to be hers, entirely at her command. Or did it have some little life of her own? She bent her finger and straightened it. The mystery was in the instant before it moved, the dividing moment between not moving and moving, when her intention took effect. It was like a wave breaking. If she could only find herself at the crest, she thought, she might find the secret of herself, that part of her that was really in charge. (McEwan 2001, p. 33)

On this picture it ought to be possible, at least in principle, to track the moment at which a person's will forms and then to identify in the activity of the brain the process that the formation of will initiates en route to the willed action. And on this picture, equally, it ought to be possible to get someone to will or not to will a certain response, and to identify in the brain the way things switch as the will comes on stream or does not come on stream. Such possibilities have been explored in recent neuroscience and related work, but the upshot is in tension with our sense of ourselves; the experiments on record throw serious doubt on the role of will, as that has been sketched so far. They suggest either that there is no such thing as agent-control or that agent-control should not be pictured on the act-of-will model. I hold by the second possibility, as I will be arguing here.

The range of experiments I have in mind is well reviewed by Daniel Wegner (2002, ch. 2), and I will just mention two varieties. Any particular

experiment can be questioned, of course, but it is hard to imagine that the observations made in these and related studies are not broadly correct; they are now part of a rapidly developing, mutually supportive body of results.

The first sort of experiment challenges the claims that when we act we have more or less infallible access to the acts of will that lie, purportedly, at the origin of our choices. Here are three fairly familiar examples:

- Subjects, with a glove on the relevant hand, are asked to draw a line on a sheet of paper in a transparent drawing box. As they draw, the line they see is actually drawn by another—this is due to an arrangement of mirrors—and begins to move away from the line they try to draw. They compensate with their own drawing and experience the gloved hand seen, and the line drawn, as under their personal control (Nielson 1963).

- Subjects are placed within an apparatus that magnetically stimulates the motor center of the brain, now on one side, now on the other. They are asked in a sequence of choices, each triggered by the sound of the magnet being activated, to raise either the right or the left index finger. Although the magnetic stimulation allows a fairly reliable prediction as to which finger will rise, subjects experience the movement as one that they control by will (Brasil-Neto et al. 1992).

- The subjects here are amputees with a phantom arm. They see an arm move where their own phantom arm, were it real, would be; this appearance is achieved with a mirror box in which the arm seen is their own non-phantom arm or someone else's arm. They experience the movement of the arm they see as a movement they themselves are controlling in the phantom arm (Ramachandran and Rogers-Ramachandran 1996).

The second variety of experiment challenges the claim that acts of will lie at the origin of what we do voluntarily and is particularly associated with the work of Benjamin Libet and his colleagues (Libet et al. 1983). Prior to that work it had been established that between 800 and 500 milliseconds before voluntary action some characteristic brain activity occurs. This came to be described as the formation of a readiness potential for action, and it was identified by many with the act of will. John Eccles, a Nobel laureate, spoke of the latency as "the mental act that we call willing" (Wegner 2002, p. 52). But in a series of experiments, Libet presented evidence that people only experience themselves as willing an action—an action they are asked to perform—about 200 milliseconds before the action itself: well after the appearance of the readiness potential. He drew a conclusion in radical disagreement with Eccles: "the initiation of the voluntary act appears to be an unconscious cerebral process. Clearly, free will or free

choice of whether to act now could not be the initiating agent, contrary to one widely held view" (Wegner 2002, p. 54).

These results establish that if willing is always needed to put an agent in control of action, and the act of will is subject to conscious experience and activation, then it is an illusion. We agents might have a generally reliable, if fallible, sense of whether something we do is a voluntary exercise. But it cannot be a general axiom, as under the act-of-will story, that an action is a voluntary exercise and is consciously a voluntary exercise, in virtue of the presence of an act of will at its origin. The first sort of experiment suggests that we identify an action as voluntary, rightly or wrongly, on the basis of something other than access to such an unmissable, unmistakable element. And the second sort shows that even if there is such an accessible element present, it comes too late in the process to do the required causal work.

These observations undermine the story that acts of will are necessary in the initiation of voluntary action and that they make the voluntary character of actions manifest and unmistakable for their agents. One likely reaction to the observations will be to think that agent-control is therefore an illusion too, since the traditional model identifies agent-control with control by acts of will. But I think that this is not an inescapable conclusion and that it is possible to remain upbeat about our capacity for agent-control.

A Response Strategy

The challenge presented is typical of a way in which science often challenges common sense: a way in which the recondite scientific image, as Wilfred Sellars (1997) puts it, challenges the image that is manifest to us in our ordinary life. A simple example of the challenge was presented by the British physicist, A. S. Eddington, when he argued that science shows that common sense is wrong even in thinking that some things are solid, others not; physics shows that everything is composed of atoms and that atoms are composed mainly of empty space.

The obvious reaction to Eddington suggests a way in which we might also react to the neuroscientific challenge. He assumed that solidity according to common sense requires that solid things be solid the whole way down, with every part being solid in the same sense as the whole. And so he argued that science shows that solidity is an illusion when it demonstrates that this connotation of solidity is unsatisfied. The obvious reaction among those of us who continue to think that 'solid' serves a useful func-

tion in commonsense discussion is to deny that this is truly a connotation of the idea or, if we think it is, to revise the notion of solidity so that it ceases to be a connotation. We may say that solidity is a property of bodies in virtue of which two or more such bodies compete for the occupation of space, and we may then argue that subatomic physics does not eliminate that property but rather serves to explain it.

When neuroscientific findings are invoked to challenge our idea of agent-control, then that is because of an assumption, parallel to that which Eddington made, that according to commonsense agent-control requires a history of action-production in which the formation of a consciously accessible will plays an initiating, causal role. We can escape Eddington's challenge by thinking of solidity without any connotation that solid things are solid the whole way down. Can we avoid the neuroscientific challenge by thinking of agent-control without a connotation to the effect that will is present in the etiology of an agent-controlled action? Can we develop a plausible conception of agent-control or freedom under which the important feature is something other than the presence of an act of will in the history of the action? I think we can.

The act-of-will story supposes that control belongs in the first place to agents and only in the second to agents. Actions display control so far as they are produced by acts of will and agents display control in virtue of their actions displaying control. But an alternative to the act-of-will picture is made salient by the observation that perhaps things are the other way around. Perhaps agents display control in virtue of having a relevant capacity for control and perhaps actions display control so far as they are performed in the domain or scope of such an agential capacity. The act-of-will story suggests that in every action I produce as an agent, I am in the loop, given a place there by my generative act of will. This alternative, in a phrase from Daniel Dennett (2003), would suggest that far from being in the loop, I am the loop. My character and capacity as an agent is what gives me control over my actions, not the particular histories of their generation. I now proceed to explore such an agent-centered view of agent-control.

If a horse runs a mile in one minute, we could say that it has the capacity to do this—the performance was characteristic of the animal—or we could say that it doesn't: the fact that it ran a one-minute mile was a fluke, not a manifestation of a standing ability. Or if the horse fails to run a mile in one minute, we could say that this shows its lack of capacity or we could hold that it has that capacity but its failing to run the one-minute mile was a contingent, not a characteristic, failure. The horse with the relevant capacity, then, could or could not run a mile in a minute, for it could or could

and consent?

agent-centered
B-children

agent-centered
view of CI

not manifest or display its capacity. Given that the capacity is real, the horse runs on each occasion in the presence of the capacity—its performance is within the domain of the capacity—but on the one occasion it manifests the capacity, on the other it fails to do so.

We can carry over this logic to the area of agent-control, if we seek to identify agent-control with an agential capacity. The control of an agent in a given domain will be constituted by a standing capacity, and the actions of the agent that are performed in the presence of that capacity will be agent-controlled. Those actions will divide, then, into two kinds. First, those that give evidence of the capacity that is present, being characteristic of the agent-controlled agent; and second, those that fail to give evidence of the capacity: those that are performed in the presence of the capacity but that fail to show what the agent is capable of.

But what capacity, if any, has a claim to be relevant to agent-control in this manner? In answer to this question, I suggest that we should look to ordinary human practice, and to the sort of capacity that we expect to find in those, including ourselves, whom we treat as agents who enjoy control.

The locus where we most explicitly display our assumption that people enjoy control is in ordinary, noncoercive conversation (Pettit 2001a; Pettit and Smith 1996). Such conversation presupposes, and continually replenishes, a currency of reasons for thought and action that are recognized as relevant on all sides, even if they are sometimes weighted differently; they are considerations that are recognized as legitimating or requiring various responses. Participants to a conversation treat one another as agent-controlled to the extent that they assume of each other—and indeed of themselves—that they have a capacity to track and conform to the demands of such reasons. The demands involved will range from requirements of respect for evidence, to requirements of consistency and valid argument, to requirements of fidelity to avowals and promises.

Imagine I authorize you as an interlocutor, treating you as someone with whom I can do conversational business. I will take you to have a capacity to think with recognizable reason—reason recognizable to you—about what is the case and to act with recognizable reason on the basis of how you think. By my lights you will be the sort of creature who is generally susceptible to reasons, and to the perception of what reasons demand: in particular, to the sort of perception that I think I can elicit in you by testimony or argument or persuasion. If I did not impute this capacity to you, I would not find you worth talking to: I might as well be talking to the wall.

I call the capacity I ascribe to you in taking this view 'conversability'. Some agents may lack conversability entirely, being out of their minds or beyond the reach of reason. And all agents are liable to lack it in some domains or in some circumstances; they will be beyond the reach of reasons on certain restricted subjects, or under certain temporary pressures. But we assume that most people have the capacity associated with conversability most of the time; that is why we think they are worth talking to. And we have no option but to assume that we ourselves generally have such a capacity, even if temporary distraction or temptation or fatigue can get in the way of its exercise. We could not sincerely put ourselves forward as candidates for conversational address by others if we did not make this assumption about ourselves. And, more important still, we could not carry on the sort of internal dialogue in which we debate with ourselves as to what we ought to think or do; we could not represent ourselves as worthy of such self-address.

I suggest that we ought to identify the capacity for agent-control with the capacity to go with recognizable reasons, in awareness of what reasons are; we might think of it, more specifically, as ^{orthonomy} the ability to be ruled by the 'orthos' or the right (Watson 2005; Pettit and Smith 1996). The suggestion fits with a now well-established tradition of discussion about free will, inaugurated by Peter Strawson's 1962 essay 'Freedom and Resentment' (Watson 1982, 2005; Wallace 1996; Wolf 1990; Dennett 1984).

If we go along with the suggestion, then we will say that to be agent-controlled in any range of action is to be conversable or orthonomous in those actions, acting in the presence of the relevant capacity: remaining, as one acts, within the reach of reason. And we can acknowledge that people who are agent-controlled in this sense may or may not always manifest conversability; they may act against recognizable reason, while retaining the orthonomous capacity to act in accord with it. We acknowledge this possibility routinely, holding people to higher expectations than their flawed performance satisfies. We recall them in such cases to the standards to which they represent themselves as faithful when they purport to be worthy interlocutors; we remind them, if you like, that they are capable of better: that, at least when exceptional cases are put aside (Frankfurt 1969), there is a sense in which they could have done otherwise than they actually did.

The capacity associated with conversability is one that we naturally conceive in a realistic, down-to-earth way. We freely admit that it comes in degrees, with each of us having bad days, bad topics, and facing hurdles of

orthonomy

varying difficulty as we seek to prove conversable. We freely admit that it is often exercised in the manner of an editor rather than an author, a conductor rather than a composer, with autonomously emerging behaviors—for example, the words that naturally come to us as we speak—being shaped to appropriate effect (Pettit 2001a, ch. 7). And we freely assume, perhaps most important, that people can have and exercise the partial, editorial control that the capacity gives them even when they are acting more or less habitually and unthinkingly. This is possible so far as the discipline of conversable reason is in 'virtual' control of action (Pettit 2001a, ch. 1). Although unthinking habit shapes what agents do, the discipline of reason will be in virtual control so far as it is ready to be activated and take charge in the event of habit failing to keep the agent in line. In that event, at least in general, the "red lights" will go on and ensure that the agent remains faithful to the perceived demands of reason (see Thalos, chapter 8 in this volume).

Beyond the Challenge

The points I have made so far are these:

- According to neuroscientific findings, agents can make mistakes about whether an action is willed, and even when agents correctly see an action as willed, they do so later than when the brain launches the action. These findings challenge the commonsense intuition that we agents have a distinctive sort of control over our actions.
- But they pose this challenge, only on the assumption that if an action is agent-controlled, then its etiology includes the formation of an act of will, where this sort of act is necessarily accessible to the agent and can be formed on demand.
- That assumption is not compulsory. In principle, we might think that an action is agent-controlled, not in virtue of the elements in its particular etiology but in virtue of the nature or constitution of the agent in whom it is produced.
- There is a good case for thinking that in practice we identify actions as agent-controlled in virtue of identifying their agents as having a suitable constitution. We do this when we identify agents as conversable or orthonomous, operating within the reach of conversationally recognizable reason.

We are now in a position to connect up these points. Suppose that we go along with the picture of agent-control as a feature of agents in the first place—their conversability or orthonomy—and a feature of actions in the

second. We can then hold that actions are agent-controlled, not because of their particular neural etiologies but because the agents are neurally so constituted that they are conversable in relation to the actions. That means in turn that we can reject the assumption made by those who pose the neuroscientific challenge. So we can escape that challenge; we can accept the findings given and still maintain a belief in agent-control.

The first sort of finding was that agents are far from infallible in identifying actions as willed or unwilled. This is not surprising under the picture presented. Assume that we learn to become relatively conversable or orthonomous over a certain range of action in the course of normal development. With any action performed within that range, we must be able to see where it is going—where our goal lies—since otherwise we would not be able to submit it to reason-based assessment. But the required ability to see where the action is going—to see, if you like, what we want or will to do—does not have to involve immediate access to an element in the etiology of the action, as under the rival picture. On the contrary, it may be an ability that is only exercised fallibly, as the action more or less gradually evolves. It may depend on a variety of cues from muscular sensation and sensory feedback, for example, not on immediate access to a presumptive act of will. And if it depends on such cues, it will be subject to precisely the sorts of error that are elicited in experiments of the first variety.

The second sort of neuroscientific finding was that agents become aware of actions as willed at some time later than when the brain begins to launch the actions. Again, this should come as no surprise under the picture presented here. What makes an action one in which I have agent-control is the fact that in this area I can be treated successfully, whether by myself or others, as conversable and orthonomous. That I can be treated that way is a function of how I am neurally constituted: a function, for example, of the possibilities of response and reform associated with the perception that what I am doing is contrary to the reasons I recognize; a function of the fact that I am not a hopeless case with whom there is no point in remonstrating, whether the challenger be myself or another. It should not be a scandal, therefore, that before ever I become aware of myself as initiating an action of that type, my brain will have done much of the work required to orchestrate the behavior. It would be surprising were it otherwise.

The experimental results discussed should no longer scandalize us, then, if we adopt the agent-centered view of control. Under this view of things, I will enjoy control of what I do so far as I can be effectively brought to book

on my actions—brought to the book of conversable reason. I will claim such control so far as I invite others to hold me to that book, as I do in presenting myself as a conversable interlocutor. To display and claim this sort of agent-control is perfectly consistent with not being the originary, self-transparent cause of what I do, as in the act-of-will picture. It only requires that however complex and opaque the process in which actions are produced, my makeup is such that the process remains sensitive to the factors that must be in place for conversability.

How should we think of decision-making under this shift of perspective? I imagine agents responding to various aspects of their surroundings, aware of prompts that would take them in one direction or another, though not perhaps aware of being aware of them. I imagine the prompts congealing to elicit a given course of action, still without the agent being aware of this activating cascade of cues. And then, as the process unfolds, I imagine an awareness of the prompt as a prompt evolving in agents, giving them a sense of where they are behaviorally headed, and taking them into the space where they can exercise the capacity that makes them conversable.

Agents may or may not be able to inhibit or reinforce the unfolding action at this point of awareness. Even if they have no such ability, however, their action will be agent-controlled, so far as they can review it in the light of the reasons available overall. They can endorse or disendorse their action, and in the case of disendorsing it, they can set themselves more or less successfully against doing the same sort of thing in the future. This is what ensures that the action is performed within the domain where conversability or orthonomy rules.

A General Perspective

The central idea in the notion of agent-control is that I am present in the performance of an action. It is not just something that happens in or to my body, unlike a spasm or reflex or compulsion. The agent-controlled action, we all think, is me; it carries my first-person signature.

The standard picture of the will in action casts the self or the 'I' as the unit of production. And in that role it represents the self as condensing—magically, as I think—in the formation of will. This act of the will, as we saw, is meant to be an event in the natural etiology of action but an event with very special properties. Although it is not itself something I do—it cannot be, on pain of infinite regress—it is a manifestation or epiphany of my self. Where it materializes, I am. And where it materializes, I can consciously be, standing in full awareness at the well-spring of behavior.

The story I prefer keeps the self in the picture, and represents voluntary action as carrying my signature. But according to this story the self is not the *unit of production* in action so much as the *unit of accountability*. Although the agent-controlled action is produced by a neural complex to which I have limited access, it remains something for which I am able to assume responsibility in the forum of exchange with myself and others. As a conversational participant, actual or potential, I purport to be someone who can generally speak for what he believes and desires; can elicit expectations about how he will behave; and can give others good reason, on pain of accepting censure or penalty, to hold him to those expectations. What it means to say that the action is agent-controlled is simply that it falls within the sphere where I prove myself to be the center of responsibility—the conversable interlocutor—that I purport to be. My agent-control has little to do with the production of action, and much to do with how effectively and authoritatively I can speak for it. This control will put requirements on the way in which the action is generated, of course—it will rule out hypnotic inducement for example—but it will not be defined by the nature of that process.

While the picture I have defended fits in many ways with our manifest sense of ourselves, as well as fitting with what we know from neuroscience and related work, it has one surprising implication (see Greene and Cohen 2004). A capacity like the capacity to be conversable or orthonomous is inevitably the product, not just of native makeup, but also of cultural development. We are not born responsible, any more than we are born free. We have to learn what responsibility requires and how to display it (McGeer 1996); as criminologists say, we have to be 'responsibilized' (Garland 2001).

Because conversability is subject to this sort of formation, it is almost bound to come in degrees. One person may be conversable over a larger or smaller domain than others, as we saw, and one person may be conversable in a higher or lower degree within the domain where the capacity is exercised. This means that whether or not people exercise agent-control in an action is never a black or white question. At this point our picture of the self in action breaks quite cleanly with the act-of-will picture. On that picture, there is bound to be a fact of the matter as to whether or not the agent's will was present in a piece of behavior; there is a bright line that will divide voluntary and properly responsible behavior from behavior that is not like this. Under our picture, by contrast, it may often be relatively indeterminate whether or not the agent was properly or fully conversable in the domain of the action; the bright line will fade into a blurry

area. But this implication of indeterminacy is surely quite plausible, however troubling it may be for moral and legal accounting. We are creatures composed out of neurally organized matter, as I said at the beginning. It would be a miracle if we turned out to display firmer contours of control than such plastic material is capable of sustaining.

The Lesson from Neuroscience

I have used the neuroscientific challenge—and, in effect, the associated findings—as a ground for arguing in favor of the view that links agent-control with conversability or orthonomy. But is there any more pointed lesson that neuroscience teaches us in this area? I believe there is. Neuroscience has taught us in many domains that what presents itself in common sense as a controlling factor in mental performance is often a side effect of a deeper, more inaccessible center of control. And it teaches precisely the same sort of lesson here.

Consider the findings of a recent study in which subjects are asked to gamble with four decks of cards, where two of the decks are stacked against them (Bechara et al. 1997). While the subjects do eventually come to register those two decks as stacked, and so to resist them, it turns out that the resistance response—or at least the disposition toward that response—is present long before any perception of the decks as stacked, or even as suspect, emerges. This is evident from fMRI imaging of what is going on in their brains when they are dealing with those decks, as distinct from the fair ones, and from associated skin conductance responses. Unconscious resistance materializes in these subjects on the basis of a subpersonal registering of the fact that things are going wrong with the suspect decks. And the eventual registering of the decks as suspect—‘there’s something I don’t like about them’—takes the form of registering them simply as decks that occasion that resistance.

The lesson of this discovery is that what appears in common sense as the factor that controls one’s response—the suspect look of the stacked decks—is not really a controlling factor but rather something that becomes available in virtue of the subpersonal controls that are operative. I do not resist the stacked decks because they look suspect, as seems to be the case from the first-person point of view; I see them as suspect because, at a deeper level, I resist them. The perception of the decks as suspect may reinforce my resistance, of course, but it is not the originating cause of resistance; that cause operates below the level of my awareness and before any awareness ever comes on stream.

What is true in this case is true in many others. I have argued elsewhere, for example, that it is not because we see something as red that we can sort it out visually and track it across different backgrounds, under different lights, from different perspectives. Rather, it is because we are equipped at a subpersonal level to do that visual sorting and tracking that we see the object as red; the red look is not a controlling factor in my mental performance but something that becomes available in virtue of the operation of lower level controls (Pettit 2003). The brain often works behind the back of the mind, and we minds—we brain-bearers—are very likely to miss the fact (Gazzaniga et al. 1998).

The picture of agent-control that we have been led to adopt in order to avoid the neuroscientific challenge bears out this general lesson. Assume that if an unfolding action is perceived by me as subject to my control then it will present itself as subject to a control I perceive myself to have; as subject, so we can put it without equivocation, to my perceived or phenomenal control. On the act-of-will picture, it is in virtue of the fact that an unfolding action is subject to my perceived or phenomenal control that it counts as agent-controlled. But, if the argument here is correct, it is not in virtue of being subject to that phenomenal control that the action is agent-controlled. Rather, it is in virtue of being agent-controlled—in virtue of being performed in the presence of a neurally supported capacity for conversability or orthonomy—that it has such a perceptual or phenomenal profile.

I see an object as red so far as I am able to sort and track it appropriately—I leave out some complications—and not the other way around. In the same manner, I perceive an action as conforming to my perceived will, so far as I perform it in the presence of a capacity for agent-control, and not the other way around. That an action is under my agent-control is the work of a vast orchestration of neural factors: those that shape the behavior but make me at the same time into a conversable subject in the domain of the behavior. When, in the absence of deceptive setups, I perceive an action as agent-controlled—when it presents itself as subject to my perceived control—this is a case of registering a sort of control that is already assured. It is not a case of assuming control for the first time.

Acknowledgment

I benefited from comments received from Walter Sinnott-Armstrong and Victoria McGeer. I am also very grateful for discussions at a University of

Alabama workshop about related matters, for responses received at a conference on the unity of science in Venice, and for a vibrant interdisciplinary discussion at a meeting of Old Dominion fellows in Princeton University.

References

- Bechara, A., H. Damasio, D. Tranel, and A. R. Damasio. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293–95.
- Brasil-Neto, J. P., A. Pascual-Leone, J. Valls-Solé, L. G. Cohen, and M. Hallett. 1992. Focal transcranial magnetic stimulation and response bias in a forced choice task. *Journal of Neurology, Neurosurgery, and Psychiatry* 55: 964–66.
- Dennett, D. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.
- Dennett, D. 2003. *Freedom Evolves*. New York, Viking.
- Frankfurt, H. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 66: 828–39.
- Gazzaniga, M., R. Ivry, and G. Mangun. 1998. *Cognitive Neuroscience*. New York: Norton.
- Garland, D. 2001. *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago: University of Chicago Press.
- Greene, J., and J. Cohen. 2004. For the law, neuroscience changes everything and nothing. *Philosophical Transactions, Royal Society* 359: 1775–85.
- Libet, B., C. A. Gleason, E. W. Wright, and D. K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity. *Brain* 106: 623–42.
- McEwan, I. 2001. *Atonement*. New York: Anchor.
- McGeer, V. 1996. Is “self-knowledge” an empirical problem? Renegotiating the space of philosophical explanation. *Journal of Philosophy* 93: 483–515.
- Nielson, T. I. 1963. Volition: A new experimental approach. *Journal of Psychology* 4: 215–30.
- Pettit, P. 2001a. *A Theory of Freedom: From the Psychology to the Politics of Agency*. Cambridge and New York: Polity and Oxford University Press.
- Pettit, P. 2001b. The capacity to have done otherwise. In P. Cane and J. Gardner, ed., *Relating to Responsibility: Essays in Honour of Tony Honore on his 80th Birthday*, pp. 21–35. Oxford: Hart.
- Pettit, P. 2003. Looks as powers. *Philosophical Issues* 13: 221–52.

- Pettit, P., and M. Smith. 1996. Freedom in belief and desire. *Journal of Philosophy* 93: 429–49.
- Ramachandran, V. S., and S. Rogers-Ramachandran. 1996. Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society* 263: 377–86.
- Sellars, W. 1997. *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press.
- Strawson, P. 1982. Freedom and resentment. In G. Watson, ed., *Free Will*, pp. 59–80. Oxford: Oxford University Press.
- Wallace, R. J. 1996. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Watson, G. 2005. *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press.
- Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Wolf, S. 1990. *Freedom within Reason*. Oxford: Oxford University Press.