Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

International Journal of Law and Psychiatry 27 (2004) 459–470

INTERNATIONAL JOURNAL OF LAW AND PSYCHIATRY

A will of one's own: Consciousness, control, and character

Neil Levy (Research Fellow)[a,*], Tim Bayne (Lecturer)[b]

[a]*Research Fellow, Centre for Applied Philosophy and Public Ethics, University of Melbourne, Parkville, Melbourne 3010, Australia*

[b]*Lecturer, Department of Philosophy, Macquarie University, NSW 2019, Australia*

1. Introduction

Talk of 'the will' has something of an anachronistic air about it today. Many people believe that the concept of the will is a holdover from an earlier, prescientific view of the human being: Before we began to unravel the connections between external stimuli, brain impulses, and action, we needed to invoke the will to explain human behavior, but reference to this mysterious entity is no longer necessary now that the cognitive sciences are deconstructing the ghost in the machine. As Daniel Wegner puts it in the title of his recent book, many think that the conscious will is an *illusion*; (Wegner, 2002).

We do not share Wegner's view. We think that reference to the will is perfectly legitimate: Certain neurobiological and psychological conditions are exculpatory precisely because they involve pathologies of the will. A full understanding of the *reactive stance*—our practice of applying what Strawson (1962) called "the reactive attitudes," such as praise and blame—must be informed by a model of the will and its pathologies. Indeed, a full account of *legal* responsibility must be informed by a model of the will and its pathologies. This article is a contribution toward such a model.

2. Contours of the will

A common methodology for delineating the contours and understanding the function of any behavioral mechanism is through an examination of the pathologies to which it is susceptible. In this section, we survey some of the more notable disturbances of the will, the better to establish its existence, and begin to appreciate its role in human action.

* Corresponding author.

E-mail addresses: nllevy@unimelb.edu.au (N. Levy (Research Fellow)), tbayne@scmp.mq.edu.au (T. Bayne (Lecturer)).

0160-2527/$ - see front matter
doi:10.1016/j.ijlp.2004.06.004

We begin with the pathology of utilization behavior. Patients with this disorder respond directly to the affordances of objects placed in front of them. Confronted with a pair of spectacles, the patient puts them on. A second pair produces the same response, although the patient is still wearing the first. Place a glass of water in front of him and he drinks, and so on (Estlinger, Warner, Grattan, & Easton, 1991; Lhermitte, Pillon, & Serdaru, 1986). Imitation behavior is a social form of utilization behavior. Patients with this condition will imitate an examiner's movements even when told not to and given negative reinforcement (Lhermitte, 1983).

It is natural to describe the actions of patients with imitation and utilization behavior as unwilled. What we mean by this description here is that the actions are elicited by environmental cues rather than being grounded in and directed by the agent's plans. The actions are indeed actions—they are not mere bodily movements; they are intentionally controlled and goal directed - but they are not grounded in the agent's standing goals. Normal agency is also elicited by environmental cures and affordances—an outstretched hand will usually prompt a handshake; but in normal human beings, such responses to the environment are modulated and inhibited by the frontal lobes, which ensure that the patient's behavior is plan-driven but environmentally responsive. Frontal lobe damage disrupts the balance between plan-driven agency and environmental responsiveness, and the patient is, as Lhermitte (1983) puts it, "subject" to all external stimuli.

Utilization and imitation behavior are often grouped with a third phenomenon—the anarchic hand syndrome - under the heading of "motor release phenomena" (Archibald, Mateer, & Kerns, 2001). Patients with anarchic hand syndrome find themselves unable to exert any (direct) volitional power over their "anarchic" hand (Della Sala, Marchetti, & Spinnler, 1991; Goldberg & Bloom, 1990). The hand may pick food up off a neighbouring plate and put it in the patient's mouth, pick up and force the patient to drink a cup of tea that the patient knows is too hot to drink, or attempt to turn the steering wheel of a car in an unwanted direction. The patient experiences no control over their hand; indeed, in some sense, they have no control over the hand, despite the fact that it is theirs.[1] Sufferers sometimes attempt to restrain their bad hand with their good hand or by sitting on it.

All three motor release phenomena involve complex, goal-directed actions that are not guided by the agent's action plans, but there is an important respect in which the anarchic hand syndrome differs from utilization and imitation behavior. The difference concerns the phenomenology of agency. Like Dr. Strangelove in the film that bears his name, anarchic hand patients are often tempted to think of their anarchic hand as possessed and controlled by an alien force. Although it is difficult to know what the phenomenology of utilization and imitation behaviors involves, these disorders do not seem to involve a substantial disruption to the normal components of the phenomenology of first-person agency. Unlike individuals with an anarchic limb, individuals who exhibit utilization and imitation behaviors do not report any peculiarities in their experience of agency (that we know of).

Motor release phenomena are not the only disorders of agency that involve pathologies of the will. Patients with Parkinson's disease also experience failures of the will. In normal agency, the agent's planner delegates the planned intention to subpersonal mechanisms that are responsible for the implementation of detailed motor routines. The planner issues the command "walk over there" in accordance with the agent's action plans, and subpersonal mechanisms activate the relevant muscles, such that one walks in the appropriate direction. In Parkinson's disease, the agent's planner appears to have become disconnected

[1] A similar phenomenon occurs in split-brain patients (see Sperry, 1968).

from the relevant subpersonal mechanisms. The patient wants to walk in a certain direction, but is unable to put the plan into action (Frith, 1992; Jahanshahi & Frith, 1998; Spence, 2001).

An additional notion of willed agency involves the notion of self-control or effort. In a series of studies, Baumeister and his colleagues have shown that people perform worse in a second self-control task undertaken shortly after a first (different) such task, and worse than controls do given a tiring task that does not require self-control, followed by a self-control task (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven, Tice, & Baumeister, 1998). Baumeister calls this phenomenon "ego-depletion": Self-control, apparently, is 'used up,' and must be recharged before it is available again. And as with muscle strength, regular exercise of self-control over the longer term can build up its resources. Ego depletion is depletion of the will; it exhausts the resources that agents must utilize to bring to a successful completion any task requiring mental exertion. It may often be ego depletion that accounts for the weakness in everyday episodes of lack of willpower, seen in such phenomena as weakness of the will, anomie, and laziness.

There is also a range of clinical phenomena that may best be understood in terms of ego depletion. Sufferers from obsessive compulsive disorder (OCD) and Tourette's syndrome report an irresistible urge to perform the associated action, but the urge does not directly cause the movement. Instead, it causes increasing levels of discomfort that can be relieved only by giving in to it. There is growing evidence that the same neurological mechanisms underlie both syndromes (Bliss, 1980; Schwartz & Begley, 2002; State, Pauls, & Leckman, 2001).

In these contexts, reference to the will is to what we might call *effort*. Whereas the patients with (say) an anarchic hand cannot control the behavior of their hand, the person whose "ego" has been depleted can, in some sense, impose their will on their actions—they just need to try hard enough. The will here is a mental muscle that one calls on to translate one's intentions into action.

Hence, there are (at least) three senses in which one might describe an action as willed (or unwilled). First, one might be referring to the genesis of the action: Was it rooted in the agent's plans, or was it generated by environmental affordances that triggered overlearned motor routines? Second, one might be referring to the phenomenology of agency: Was the action accompanied by the experience of doing, or did the agent experience the action as alien and unowned? Third, one might be referring to the degree of effort involved in prosecuting the action: Did the action demand some degree of self-control, or was it performed effortlessly?

Exactly how these three senses of the will are related is an open question. Must actions that are willed in one sense of the term also be willed in the other senses of the term? And if so, why? Is one of the three notions of willed agency that we have identified more fundamental than the other two, or do we have a cluster of related but relatively autonomous notions? Addressing these questions in depth would demand an article in its own right, but we can make some tentative observations here.

The exercise of self-control seems to be parasitic on plan-based control and the phenomenology of first-person agency being in place. In exercising self-control, one attempts to make it the case that one's actions conform to one's (reflectively endorsed) action plans. And the process of exercising self-control seems to assume the phenomenology of first-person agency: It seems likely that a person with no sense of being an agent would be unable to exercise any degree of self-control.

Must actions that derive from one's action plans involve self-control? It seems not. Many of our everyday behaviors—driving a car, waving goodbye, typing a paper—count as willed, in the sense that there are governed (at least distally) by our action plans, but they involve little in the way of self-control or effort. And actions can be governed by action plans (again, at least distally) without possessing much

in the way of the phenomenology of first-person agency: Experienced drivers will often change gears without experiencing themselves as engaged in such an activity. (In many cases, attending to the experience of performing an action can actually degrade performance.) No doubt, there are many important connections between these three notions of the will, but the considerations adduced here suffice to show that there is also a certain degree of independence between them.

3. Resisting will-talk

Although implicit reference to willed agency is common in the cognitive sciences, many cognitive scientists nevertheless take a sceptical view of will-talk. Will-scepticism has a number of roots, but perhaps, the dominant objection involves the worry that it is antiscientific. In recent work, Daniel Wegner sounds the note of will-scepticism:

> The fact is, it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. *Although it is sobering and ultimately accurate to call all this an illusion*, it is a mistake to conclude that the illusory is trivial. On the contrary, the illusions piled atop apparent mental causation are the building blocks of human psychology and social life. (Wegner, 2002, p. 342; emphasis added)

Wegner urges that we adopt his theory of apparent mental causation in place of the folk notion that the conscious will is causally efficacious.

> The theory of apparent mental causation turns the everyday notion of intention on its head. . .The theory says that people perceive that they are intending and that they understand behavior as intended or unintended—but they do not really intend. Instead, conscious thoughts coming to mind prior to action are described in the theory as previews of action, ideas that surface into consciousness as the result of unconscious processes like those that create action itself. (Wegner, 2003, p. 221)

Among the difficulties one faces in evaluating Wegner's claims is that it is unclear exactly what he means by the will (or, the "conscious will"). Does he mean plan-based agency? Does he mean the phenomenology of agency? Or does he mean the (experience of) effort and self-control? In short, what exactly does Wegner mean when he says that the conscious will is an illusion? Although we are not sure that Wegner has a unitary conception of the will, the passage quoted above suggests that he thinks of willed agency as intentional agency, which suggests that he has something like our plan-based conception of the will in mind.

We read Wegner as offering three lines of arguments for the claim that the will, that is, plan-based agency, is an illusion. The first argument involves an appeal to the idea that subpersonal mechanistic explanations of human behaviour displace or eliminate personal-level explanations. We experience ourselves as intending, acting, and deciding, but, says Wegner, such experiences are nonveridical: Human behavior really springs from subpersonal mechanisms.

Wegner assumes that personal and subpersonal accounts of action are in competition, but an alternative conception of the relationship between them is to see them as complementary. One might regard subpersonal explanations of behavior as explaining *how* intentional (plan-based) agency works rather than explaining it away. What Wegner's theory of apparent mental causation provides—if true—is an account of the mechanisms that are involved in generating the phenomenology of first-person agency,

what Wegner call the "feeling of doing." But to explain how the feeling of doing is generated does not, in and of itself, show that no one does anything. It is one thing to show how certain experiences are generated, it is another thing to show that they are (systematically) nonveridical.

A second strand of Wegner's case for will-scepticism involves an argument from dissociation. He attempts to show that there is a range of cases—both inside and outside the laboratory—in which agency and the experience of agency can come apart. On the one hand, one can perform actions without experiencing ourselves as performing them, and on the other hand, one can experience an action that someone else is doing as something that one is doing. Wegner argues that many apparently occult phenomena, such as table turning and the ouija board, are instances of the first sort of dissociation: The agents in question are doing things that they do not realize they are doing. Wegner's advances his 'I-Spy' laboratory game are an example of the second sort of dissociation. In this experiment, Wegner shows that if participants are primed to think about a particular image on the screen, they are likely to experience themselves as moving the cursor to that image, and hence, believe that they moved the cursor to that image, even when the experimenter rather than they themselves was responsible for moving it.

We have two reservations about Wegner's argument from dissociation. First, Wegner's case for thinking that willed agency and the experience thereof can dissociate is not quite as straightforward as it might at first seem. Consider his interpretation of "Penfield actions," that is, actions that the neurosurgeon Wilder Penfield generated by stimulating the exposed brains of conscious patients (Penfield, 1975). Upon stimulation, Penfield's patients (1975) moved their hands or vocalized, but they did not seem to experience any sense of volition with respect to their actions and claimed not to have done them willingly. Wegner presents these actions as willed actions that the patient fails to experience as willed. But it is far from clear that Penfield actions qualify as willed actions—indeed, there is some temptation to deny that Penfield actions even belong to Penfield's patients.[2] Penfield's patients experienced "their" actions as unwilled occurrences, and perhaps, that is precisely what they were. Suppose that Penfield had known the consequences of stimulating the patient at the particular spot at which he stimulated him. Now, it seems more attractive to say that Penfield generates the action. But giving Penfield more knowledge does not change the patient's relation to the action: If they do not perform it in the latter case, then they do not perform it in the former case either.

But even if Wegner's claim to have shown that experiences of willed agency can be nonveridical were true, it is hard to see how it would support the further claim that the conscious will is an *illusion*. The argument seems to be 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. But this inference is no more compelling than the inference from the fact that because visual illusions occur, we have no reason to trust our normal visual experience. Indeed, one might think that Wegner's work brings into stark relief just how reliable the mechanisms that produce the sense of agency are: Dissociations between the experience of agency and the actual exercise of agency are striking exceptions to the norm.

A third strand of Wegner's case for will-scepticism draws on Libet's famous studies on the 'readiness potential' (Haggard, Clark, & Kalogeras, 2002; see also Haggard & Eimer, 1999; Libet, 1985; Libet,

[2] Wegner (2002) seems to equivocate on the question of whether experience of conscious will is necessary for voluntary agency. On the one hand, he says that "without an experience of willing, even actions that look entirely voluntary from the outside still fall short of qualifying as truly willed" (p. 3). On the other hand, he says: "it appears possible to produce voluntary action through brain stimulation with or without an experience of conscious will" (p. 47).

1999). The readiness potential is an electrical event in the brain that is associated with readiness for action, hence, the name. Libet (1985) asked participants to move a hand 'at will' and to note when they felt the urge to move by observing the position of a pointer on a special clock. While they were doing this, Libet recorded the readiness potential of his participants. He found that it *preceded* the conscious awareness of the urge to move by about 400 ms.

Libet and others have claimed that these results constitute a difficulty for those who want to invoke the will to explain actions. Suppose that we conceive of the will as the mechanism that explains (or perhaps *constitutes*) conscious control over actions. Libet's (1985) results suggest that the brain activity that underlies action begins before the participant is consciously aware of the urge to move, which in turn, suggests that the conscious urge to move is epiphenomenal: The will can play no role in agency because it comes on the scene too late.

There are several reasons to think that Libet's (1985; 1999) results do not provide as strong a case for intentional epiphenomenalism as many think they do. First, we need to consider what is involved in the notion of a conscious decision to act. Arguably, Libet confuses the decision of which his participants became conscious with the state by means of which they became conscious of their decision. There is no reason to think that the *process* of decision making should always be conscious, in the sense that there is something it is like to make a decision. For most everyday decisions (when to pick up a pen, which hand to use to answer the telephone, etc.), we are not aware of a process of decision making. Moreover, even on those relatively rare occasions when we cannot make up our minds, we are conscious of our decision-making process, much of it still necessarily takes place below the level of conscious awareness. Assigning weight to considerations, for instance, cannot be a conscious process, not if decision making is a rational process, for if considerations only have a weight in light of our decisions, then that weight is arbitrary. The considerations on the basis of which we decide must therefore be experienced by us as already having a certain weight, independently of our decision—presumably, subpersonal mechanisms take care of this for us, in the light of our preexisting system of values and ends. Once weights have been assigned, we choose the alternative that is weightier (once more, if we decide rationally). Hence, there is a sense in which a large part of rational decision making is outside our control, for conceptual rather than neurological reasons. We need not appeal to Libet's experimental results to find this out. As Dennett (1984) has pointed out, decisions seem at once to be paradigms of voluntary agency, and 'strangely out of our control' (p. 78), at once actions we undertake and events we undergo.

Second, it is unclear whether we should identify the readiness potential with the (neural substrate of) participants' decision to raise their hand. As Mele (2003) points out, one could take the readiness potential to underlie or constitute the neural substrate of desires or urges rather than intentions or decisions. 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 experiments do suggest that consciousness plays a role in certain types of decision making. Libet's participants are conscious when he gives his instructions to them, and they have to (consciously) comply with the requests made of them (Flanagan, 1996). Arguably, Libet's participants have consciously delegated the initiation of their actions to subpersonal mechanisms. This would not be unusual: Conscious processing is in the business of passing control of the details of an action to subpersonal mechanisms. We consciously decide to ask for a coffee, but we (typically) do not consciously decide which words to use in making the request, how loudly to say them, how to pronounce them, and so on.

4. The will and responsibility

We have suggested that our conception of the will involves three (related) phenomena: plan-based control, the phenomenology of agency, and effort. In what follows, we will sketch two (interconnected) ways in which one might connect these phenomena with questions of culpability: via the notions of control and of character. We examine these two options in turn.

4.1. Control

It is very plausible to think that we are responsible only for actions over which we exercise a certain form (or degree) of control. Agents are excused blame concerning all events for which they are *causally* responsible if they were unable to exercise relevant control over their occurrence. I am excused from responsibility for causing your glass to break if I did so as a result of a sudden and unprecedented muscle spasm, whereas I am not excused (other things equal) if I was in control of my actions at the time. Control is a necessary condition for moral responsibility: Intuitively, to the degree that a person has lost control of her actions, the less appropriate it is to blame or praise her for them.

But it is not control alone, but *rational control*, that seems to be central to the reactive stance. Patients who exhibit imitation and utilization behaviour have some form of control over their actions—their movements are intentional and goal directed—but they are not under the patient's rational control. Similarly, the somnambulist who drives 23 km to his parents-in-laws' home (Broughton et al., 1994) responds to his environment in a controlled manner, but his actions are not rationally guided. In this kind of case, actions seem to be driven by environmental affordances rather than agential plans. Thus, the agent's actions are (in some sense) controlled, but they are not *rationally* controlled. We suggest that in cases like this, and in analogous cases, such as those involving epileptics who suffer a seizure while walking or driving and continue on to their destination in an apparent state of unconsciousness, agents are able to use only a subset of their relevant intentional states to guide their actions (Levy & Bayne, in press).

Disorders of control can be divided into two classes, failures of *authority* and of *inhibition*. Failures of inhibition occur when the agent in question has lost rational control over their actions, but the action in question can nonetheless be ascribed to the agent. In failures of authority, by contrast, the loss of control is such as to call into question the ascription of the action to the agent. Consider these categories in the light of the following syndromes.

In Tourette's syndrome, agents might find themselves saying aloud words that have 'crossed their mind' in the normal way. Their words express their (fleeting) thoughts, but unlike most of us, they find it extremely hard to inhibit them—so hard that it is inevitable that they give in. Their actions are expressions of their mind, but their content is divorced from their action plans and is expressed despite their efforts at inhibition. We can say that the degree to which they are willed is therefore much lower.

Utilization behaviour is somewhat similar to Tourette's syndrome in that the actions are divorced from action plans, although here, the phenomenology seems to be different and no effort is made to inhibit the action. The difference between the two syndromes suggests a further subdivision of failures of inhibition into those in which the agent is conscious that they have strong reasons against the action, which they nevertheless go on to perform, and those in which they seem unaware of any such reasons. In OCD and Tourette's syndrome, agents perform their actions for a reason (to relieve the increasing discomfort that continued resistance causes in them), and they perform them as a result of a conscious act of 'giving in';

in this sense, their actions are willed. However, OCD and Tourette's syndrome (in the case of the latter typically, but not invariably) are ego dystonic: Sufferers do not identify with the actions they perform in its grip, they perform them only to relieve discomfort. Whereas in normal action, we act to bring the world into line with our desires, and any relief of frustration or feeling of satisfaction is typically only a welcome side effect, in these syndromes, the entire point of the action for their sufferers is this relief (Schroeder, in press). Utilization behavior and imitation do not seem to be ego dystonic in this manner; sufferers do not regard themselves as having reasons to inhibit the behavior. It may be that ego depletion is best understood on the model of ego-dystonic inhibition failure.

In both kinds of failure of inhibition, there is a clear link between the agent and the action. They perform it for a reason that they are normally capable of acknowledging (e.g., because that is what glasses are for or to relieve intense discomfort), even if these are not reasons that a rational agent would regard as sufficient to motivate action. In failures of authority, if there is such a link, it is not one that the agent is in a privileged position to identify. Such failures—anarchic hand is the paradigm pathological form—occur when the agent's body moves not merely without their consciously intending it, but even in defiance of their intention to keep it still or to move it in a different way.

There is good reason to follow Peacocke (2003) here and deny that actions that involve failures of authority should be ascribed to (apparent) agents at all. Agents with an anarchic hand are typically no better placed to predict its movements than are third-person observers. In such a case, it is clear that, and how, failures of authority excuse the agent from moral responsibility. If my hand engages in a morally problematic act that I neither intended nor foresaw, my responsibility for it is greatly reduced or even dissolved (although complications set in when we consider the fact that I might have been able to exercise indirect control over my hand, and when we consider the question whether someone with an anarchic hand is capable of learning to predict its behavior).

It is far less clear how, and to what extent, failures of inhibition excuse. We think that different failures of inhibition excuse to different extents and in different situations.

Failures of inhibition can be divided into *partial* and *total* failures of inhibition. In total failures of inhibition, agents are unable to stop themselves from acting on their urge. This can happen in one of two ways: (i) either the urge is much stronger than are the resources for self-control that are available to the agent, so that these resources are overwhelmed; or (ii) the urge, or some other feature of the situation, prevents the agent from bringing their resources of self-control to bear on the situation. In either case, the agent acts compulsively (Holton & Shute, unpublished). Partial failures of inhibition are different in that the agent could always hold out against the urge a little longer (say, if the stakes were raised). If I am severely ego depleted, it is inevitable that I will soon give in to the temptation to stop the self-control task in which I am engaged. But I could, through an effort of will, continue in it a little longer.

It is clear that the distinction between total and partial failures of inhibition is important to assessments of moral responsibility. If addiction is to be understood on this model, for instance, then the addict is not responsible for giving in to the urge to consume the drug *sooner* or *later*. But they might be responsible for giving in *when* they did. If there were good reasons for them to wait a little longer, then they might be blameworthy for not holding out. The details of the case have an important bearing on the manner and extent to which the agent is excused. Partial failures of inhibition generally have a phenomenology that is highly aversive: The addict suffers withdrawal pangs, the OCD sufferer experiences mounting anxiety, and so on. Excuses, partial or total, in this kind of case can be modelled on cases of coercion. We continue to blame agents, in this kind of situation, to the extent to which the moral stakes make it reasonable to demand of them that they hold out against their suffering. Some cases

will turn not so much on the aversiveness of the phenomenology as on the effects that resistance has on the cognitive state of the agent: Some kinds of drug addiction, for instance, may cloud the capacity for clear thinking more than they produce painful withdrawal symptoms. Of course, many cases will combine cognitive and affective excusing conditions.[3]

Partial failures of inhibition therefore offer excuses, inasmuch as the agent acts under the effects of coercion and judgment-clouding conditions. Whether the excuses available in these cases are partial or total will depend upon the details of the case and the moral stakes: Agents may remain blameworthy for acting as and when they did. With total failures of inhibition, in contrast, the agent is not blameworthy for acting as and when they did (other things being equal).

4.2. Character

Some philosophers ground moral responsibility in the *character* of the agent. Some philosophers think that moral responsibility is important, at bottom, because in acting, we express ourselves, leaving a mark on the world which is distinctively ours (Fischer, 1999). On this view, failures to appreciate reasons or to conform our actions to reasons are exculpatory to the extent that the resulting actions do not reflect our character. Character is a reflection of the totality of our reasons and plans, and actions that do not derive from our plans do not (fully) reflect our characters. On this view, bad actions are bad because they reflect and stem from a bad character.

Philosophers who embrace characterological foundations for moral responsibility argue that control is important only insofar as it allows us to distinguish between actions that (fully) reflect the agent's character and those which reflect only (at most) a part of it (Reznek, 1997). However, it is possible to argue for a more subtle use of characterological notions. Rather than identify an agent's character with the mechanisms that underlie the normal control of their actions, we might take a first-person approach to character, equating it with those aspects of the self with which the agent identifies. The pioneer of this kind of approach to moral responsibility is Harry Frankfurt. Frankfurt (1987) argues that agents are fully responsible for their actions only if they are the product of desires that they endorse. This accounts for the intuitive plausibility of mitigating blame for drug addicts who struggle against their addiction, as compared with addicts who experience no conflict between their cravings and their sense of who they think they ought to be.

Cases in which agents are moved by desires with which they fail to identify their responsibility seem to be diminished, even when they meet the strict standards for control urged by Reznek (1997). Perhaps, our intuitions here are the product of the knowledge that agents will typically struggle against desires that they do not endorse. To the extent to which conditions produce ego-dystonic desires—as do OCD, drug addiction (in some cases, at least), Tourette's syndrome, and perhaps, ego depletion—agents ought to be partially excused blame for their actions because their actions do not (fully) reflect their character.

[3] If it is often possible to hold out a little longer, are these not cases in which it is possible to hold out indefinitely and, therefore, in which no excuse is available? Often, the answer is no. First, from the fact that it is possible to hold out a little longer, it does not follow that it is possible to hold out indefinitely; second, from the fact that it is possible to hold out longer—even much longer—it does not follow that there is no (perhaps partial) excuse for giving in now. Simple muscle fatigue illustrates both points. As my muscles tire, the temptation for me to stop exercising, and the difficulty of continuing, increases. Nevertheless, it is likely that I could continue for some time after the point at which I stop; to that extent, my stopping is my choice. But it is not true that I could continue to exercise indefinitely; eventually, I would literally collapse from exhaustion (and it may be that prior to this point I would no longer exercise *rational* control because my fatigue would affect my mental processes). Moreover, although I normally choose the point at which I stop exercising, if there is any blame attached to my choice, it is usually mitigated by the knowledge that I acted under the coercive effects of exhaustion.

One attractive feature of the characterological account of moral responsibility is that it seems resistant to future Libet-style objections to free will. We saw that there are good reasons to resist Libet's claim that consciousness does not play a very significant role in the initiation of action. But it may be that some surprises are in store for us. Perhaps, neuroscientists will one day succeed in demonstrating that conscious reasons-responsiveness plays little or no role in the production of behavior. What then? It is here that character-based conceptions of responsibility appear most attractive. My actions may reflect my character even when they are not under my direct conscious control, and for this reason, I might be held responsible for what I do. If, for instance, my actions remain reasons-responsive, although the mechanisms of reasons-responsiveness operate below the level of conscious awareness, moral responsibility would remain a respectable notion.

5. Responsibility for the will

My failure to stick to my diet or exercise regime might be taken to indicate a lack of willpower on my part. More seriously, we may talk about the immense effort of will needed to overcome a drug addiction. Some people may blame drug addicts for their addiction and, derivatively, for the actions prompted by it, claiming that lack of willpower is itself blamable. Others claim instead that agents are responsible only for what they *can* will, and that therefore, a lack of will—a lack of mental muscle—is exculpatory. A person can be held accountable for failing to put more effort into a task than they did, but their failings are excused if the amount of effort required was more than they were capable of providing. Of course, in some situations, one might be culpable for not having developed one's capacities of self-control, but even taking this point into account, it will remain true that some individuals have less will power than others do through no fault of their own.

Intuitively, what we need here is an account of the difference between the capacities of self-control and the exercise of those capacities. We need an account of when a person had the self-control necessary to resist a certain type of provocation but failed to draw on their resources of self-control, and when they failed to have the resources of self-control they needed to resist the provocation. But how are resources of self-control to be measured? We can perhaps measure how much self-control someone actually exhibits on particular occasions, but it is less clear how one could measure how much control they could have exhibited on those occasions. We leave this as an outstanding problem.[4]

6. Conclusion

Far from casting doubt on the notion of the will, the study of pathologies of human behavior provides powerful reasons for insisting on its relevance to understanding moral agency. We cannot understand what goes wrong in the pathologies of the will unless we postulate that normal human action is under the rational control of agents. Just as rationality is undermined in delusional mental illness, so too willed agency is undermined by utilization behaviour, the anarchic hand, ego depletion, and related phenomena. But as we have seen, these phenomena undermine willed agency in different ways. A full model of the

[4] Smith (2003) provides some of the elements for the required account of agent capacity.

reactive stance stands to benefit from the study of the various facets of the will and the pathologies to which they are subject.

Acknowledgements

The authors gratefully acknowledge the assistance of the Australia Research Council Discovery Grant DP0452631 in funding this research.

References

Archibald, S. J., Mateer, C. A., & Kerns, K. A. (2001). Utilization behavior: Clinical manifestations and neurological mechanisms. *Neuropsychology Review*, *11*(3), 117–130.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego-depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Bliss, J. (1980). Sensory experience of Gilles de la Tourette syndrome. *Archives of General Psychiatry*, *37*, 1343–1347.

Broughton, R., Billings, R., Cartwright, R., Doucette, D., Edmeads, J., Edwardh, M., et al. (1994). Homicidal somnambulism: A case report. *Sleep*, *17*, 253–264.

Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, *29*(11), 1113–1127.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Estlinger, P. J., Warner, G. C., Grattan, L. M., & Easton, J. D. (1991). Frontal lobe utilization behavior associated with paramedian thalamic infarction. *Neurology*, *41*, 450–452.

Fischer, J. M. (1999). Responsibility and self-expression. *The Journal of Ethics*, *3*, 277–297.

Flanagan, O. (1996). Neuroscience, agency, and the meaning of life. In *Self-Expressions* (pp. 53–64). Oxford, UK: Oxford University Press.

Frankfurt, H. (1987). Freedom of the will and the concept of a person. In Frankfurt (Ed.), *The importance of what we care about* (pp. 11–25). Cambridge, UK: Cambridge University Press.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove, UK: Lawrence Erlbaum Associates.

Goldberg, G., & Bloom, K. K. (1990). The alien hand sign. Localization, lateralization and recovery. *American Journal of Physical Medicine & Rehabilitation*, *69*, 228–238.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*, 382–385.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and awareness of voluntary movements. *Experimental Brain Research*, *126*, 128–133.

Holton, R., & Shute, S. (unpublished). *You can't lose what you ain't never had: Self-control in the modern provocation defence*.

Jahanshahi, M., & Frith, C. (1998). Willed action and its impairments. *Cognitive Neuropsychology*, *15*, 483–533.

Levy, N., & Bayne, T. (in Press). Doing without deliberation: Automatism, automaticity, and moral accountability, *International Review of Psychiatry*.

Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–255.

Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes: Part I. Imitation and utilization behavior: A neuropsychological study of 75 patients. *Annals of Neurology*, *19*, 326–334.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*, 529–566.

Libet, B. (1999). Do we have free will? In B. Libet, A. Freeman, & A. Sutherland (Eds.), *The volitional brain: Towards a neuroscience of free will* (pp. 47–57). Thorverton: Imprint Academic.

Mele, A. R. (2003). *Motivation and agency*. Oxford, UK: OUP.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774–789.

Peacocke, C. (2003). Action: Awareness, ownership, and knowledge. In J. Roesssler, & N. Eilan (Eds.), *Agency and self-awareness* (pp. 94–110). Oxford, UK: Oxford University Press.

Penfield, W. (1975). *The mystery of the mind: A critical study of consciousness and the human brain*. Princeton, NJ: Princeton University Press.

Reznek, L. (1997). *Evil or ill? Justifying the insanity defence*. London: Routledge.

Schroeder, T. (in press). Moral responsibility and Tourette Syndrome. *Philosophy and Phenomenological Research.*

Schwartz, J. M., & Begley, S. (2002). *The mind and the brain*. New York: ReganBooks.

Smith, M. (2003). Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In S. Stroud, & C. Tappolet (Eds.), *Weakness of will and practical irrationality* (pp. 17–38). Oxford, UK: Clarendon Press.

Spence, S. (2001). Disorders of willed action. In P. Halligan, C. Bass, & J. Marshall (Eds.), *Contemporary approaches to the study of hysteria* (pp. 235–250). Oxford, UK: OUP.

Sperry, R. W. (1968). Hemisphere deconnection and unity in conscious awareness. *American Psychologist*, *23*, 723–733.

State, M. W., Pauls, D. L., & Leckman, J. F. (2001). Tourette's syndrome and related disorders. *Child and Adolescent Psychiatric Clinics of North America*, *10*, 317–331.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, *48*, 187–211.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: The MIT Press.

Wegner, D. (2003). The mind's self-portrait. *Annals of the New York Academy of Sciences*, *1001*, 212–225.