

The Feeling of Doing: Deconstructing the Phenomenology of Agency

Tim Bayne
University of Oxford and St. Catherine's College
Manor Road
Oxford OX1 3UJ
United Kingdom
tim.bayne@gmail.com

Dr Neil Levy
Centre for Applied Philosophy and Public Ethics
Department of Philosophy
University of Melbourne
Parkville Vic 3010
Australia
nllevy@unimelb.edu.au

This paper appeared in (2006) N. Sebanz and W. Prinz (eds.) *Disorders of Volition*.
Cambridge, MA: MIT Press, 49-68. Please consult the published version for
purposes of quotation.

1. Introduction¹

One of the most exciting developments in the cognitive sciences in recent years has been a rediscovery of the phenomenology of agency (see e.g. Graham 2004; Horgan et al. 2003; Nahmias et al. 2004). That the phenomenology of agency has received renewed attention is due in no small part to claims that it is at odds with what the cognitive sciences are revealing about the structure of agency itself: in short, that the manifest or phenomenological image of ourselves as agents is inconsistent with the image emerging from the cognitive sciences. Wegner speaks for many when he writes: "... 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 illusion 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)

One response to will-skeptics—as we shall call them—is to challenge their interpretations of the data derived from the cognitive sciences. This is a project that we have pursued in other work (Levy & Bayne, 2004), but it is not the one we will pursue here. Instead, we will concentrate on the task of examining the phenomenology of agency. We suspect that much of the motivation for the current wave of will scepticism derives from rather naïve models of the phenomenology of agency. A more nuanced account of the phenomenology of agency might fit rather better with what the cognitive sciences are telling us about ourselves.

Clarifying both the content of the phenomenology of agency and its relationship to the scientific image of ourselves as it is being revealed by the cognitive sciences is essential to a full understanding of disorders of volition, for many disorders of volition manifest themselves in the form of abnormal experiences of volition. Sometimes these abnormalities are reflections of the fact that the actual structure of agency has itself been disrupted, but in other cases it may be only that the mechanisms responsible for the phenomenology of agency are disrupted.

There are many components within the experience of first-person agency. Or, as we can also put it, the experience of first-person agency includes many other experiences as components. (Assume from now on that by ‘the phenomenology of agency’ we mean ‘the phenomenology of first-person agency’.) We will focus on just three components: the experience of mental causation; the experience of authorship; and the experience of effort.² These experiences have representational content. That is, they present the world – in this case, the agent and her actions—as being a certain way. Our main aim in this chapter is to examine the ways in which these experiences represent an agent and their actions.

2. The phenomenology of 'mental causation'

We typically experience our actions as purposive. We do not simply find ourselves walking towards a door and, on the basis of this, form the belief that we must be intending to open it; instead, we experience ourselves as walking towards the door in order to open it. This sense of goal-directedness can operate at a number of levels. For example, one might experience oneself: walking towards a door in order to open it; opening the door in order to feed the dog; and feeding the dog in order to keep him quiet. The phenomenology of a single action can include the nested purposes for which the action is being performed.

How should we understand the experience of purposiveness? Is it the central component of the phenomenology of agency? Perhaps experiencing a movement as an action *just is* to experience it as implementing an intention that one experiences as one's own. Call this the *purposiveness thesis*. The purposiveness thesis suggests that disorders in the phenomenology of agency can arise either from failures to accurately track the contents of one's intentions, or from failures to track the fact that one's intentions are one's own (Wegner 2002, Frith 2002; this volume).

An obvious objection to the purposiveness thesis is that we perform many actions without experiencing these actions as being performed on the basis of a specific intention (Marcel, 2003). To give an example of Searle's (1983), when deep in thought one might suddenly get up and start pacing around the room. Actions of this kind – what, following Bach (1978) we will call *minimal actions* – might be caused by intentions, but that is not how they are experienced. It seems possible to experience oneself as performing an action without experiencing that action as the result (or implementation) of an intention. Call this the objection from *minimal actions*.³

There are a number of ways in which a proponent of the purposiveness thesis might respond to this objection. We will mention three. First, one might deny that minimal actions are accompanied by a sense of agency. Minimal actions, on this line of thought, might be actions, but they do not feel like actions. We think the proponent of the purposive thesis would be unwise to adopt this position: minimal actions sure seem like actions to us! There seems a marked

phenomenological difference between feeling one's arm go up in the context of an automatic action and feeling one's arm being raised by a friend. The former feels like an action, the latter does not.

A second response to the objection from minimal actions is to claim that minimal actions are accompanied by an awareness of intentions. The proponent of the purposive thesis might admit that minimal actions are not accompanied by an awareness of what Searle (1983) calls *prior intentions*, but she might insist that there is an intentional component – what Searle calls an *intention in action* – to all experiences of agency. In fact, this seems to be Searle's view. Searle holds that the content of an experience of agency *just is* the content of an intention in action (Searle, 1983, 91).

We grant that there are fine-grained intentions that do not merely trigger actions but govern their evolution in a dynamic way. And we grant that we can be – and often are – aware of their role in shaping our movements. But is the presence of such intentions essential to the phenomenology of agency? Can we not experience a movement as an action without experiencing it as implementing an intention-in-action? We think so, although the elusiveness of the phenomenology of agency makes this a difficult point to confirm.

A third response on the part of the purposive theorist to the minimal actions objections is to modify the view. The original idea with which we began is that the experience of agency involves an awareness of the *content* of the intention that is implemented in the target action. A weaker, and more plausible view, is to hold that the experience of agency involves an experience of the action as having a certain intention, without necessarily involving an awareness of the content of that intention. On this view, to experience a movement as one's action involves an experience of it as implementing an intention (which one experiences as one's own). This view seems plausible, but it is a long way from the original purposiveness theory.

We turn now to a different set of issues raised by talk of the experience of purposiveness. What is involved in experiencing an action as being performed in virtue of a certain intention? What is involved in experiencing a movement as

implementing a certain intention? One view, which seems to be widely endorsed, adopts a causal approach to the phenomenology of purposiveness. According to what we call the *mental causation thesis*, experiencing oneself as opening a door involves an experience of the intention to open the door as causing one's movements.

Note two important points about the mental causation thesis. First, it is a claim about the phenomenology of agency, not the structure of agency. One can reject the mental causation thesis without rejecting the causal theory of agency itself (and vice-versa). Second, the mental causation thesis should not be confused with the claim that we experience our *experiences* of agency as causing our movements, a view that Searle (1983) advocates. We find this Searlean view implausibly strong.

Is the mental causation thesis plausible? There is room for scepticism. As Horgan et al. write: "Your phenomenology presents your own behavior to you as having yourself as its source, rather than (say) presenting your own behavior to you as having your own occurrent mental events as its source" (Horgan et al., 2003: 225). It is doubtful whether the phenomenology of purposeful agency involves an experience with causal content, at least if by "cause" one means anything more than counterfactual dependence. We do not, it seems, experience our intentions as sufficient conditions for our movements. In making these claims we are not relying on a scepticism about the phenomenology of causation; unlike Hume, we grant that causal relations can be experienced. We simply doubt that, in the usual run of things, the experience of first-person agency involves an experience of mental causation as such.

In fact, the experience of one's movements being caused by one's intentions seems to be characteristic of some disorders of volition. It is sometimes suggested that one of the pathological features of the phenomenology of addiction and obsessive-compulsive spectrum disorders is that the individuals concerned experience their actions as caused by their desires and urges rather than as having their source in *them*.⁴ We suspect that as one begins to experience one's movements as caused by one's mental states, one no longer experiences them as

one's own actions. That is, one no longer experiences the action as an instance of *first-person* agency.

A further problem for proponents of the mental causation thesis is that it seems possible to experience one's movements as being caused by one's intentions without experiencing the movements in question as actions. Consider the following variant on a classic thought-experiment from the action theory literature (Davidson, 1973):

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might lead him to form the intention to loosen his hold on the rope, and this intention might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. The climber experiences the loosening of his hold as caused by his intention to loosen his hold, but he does not experience the loosening of his hold as an action.

Davidson introduced the case of the unfortunate climber in order to demonstrate that causal accounts of action face a problem of deviant causal chains. His point was that even though the loosening of the climber's hold was caused by the climber's beliefs and desires, a causal theorist shouldn't say that the climber's loosening of his hold was an action. Causal accounts of action need to say when the causal relation between beliefs and desires (and intentions) and movement generates agency, and when it is merely deviant. Our motivation for referring to the unfortunate climber is to point out that causal accounts of the *phenomenology* of agency face a parallel problem. They need to say when the experience of a causal relation between an agent's intentions generates the experience of agency and when it does not. We will leave the task of exploring possible solutions to this problem for another occasion.

4. The phenomenology of authorship

We turn now from the experience of mental causation to the experience of ourselves as agents – what we will call the *phenomenology of authorship*. We begin by asking how the experience of authorship might be related to two other experiences mentioned in connection with the phenomenology of agency: the experience of self as source (Horgan et al 2003), and the experience of agent causation (O'Connor 1995).

Talk of an experience of 'self as source' seems to capture certain features of the experience of agency. As Horgan and co-authors say, we experience our actions as deriving from ourselves rather than deriving from our mental states. But how exactly should we understand this relation of being a source? Should we understand it in causal terms? Horgan and co-authors don't want to commit themselves to a causal reading of this notion, but it is unclear how to read it if not in causal terms.

As the term suggests, agent causal theorists adopt an explicitly causal conception of the phenomenology of being an agent. According to O'Connor, "[agent causation] is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favour doing so; it seems to be the case, rather, that I produce my own decisions in view of those reasons..." (O'Connor 1995, 196). Ginet also gives an agent causal gloss on the phenomenology of agency, although he does not himself endorse agent causation: "My impression at each moment is that *I* at that moment, and nothing prior to that moment, determine which of several open alternatives is the next sort of bodily exertion I voluntarily make" (1990, 90).

Agent causation is the thesis that agents are, or at least can be, primitive causes of their movements, where the force of the 'primitive' indicates that the causal relation between the agent and the movement cannot be reduced to relations between events (Chisholm 1982; O'Connor 1995; Taylor 1966). Proponents of agent causation typically claim that the causal relation between the agent and their movements is not necessitated by prior states of the agent. On this view, agents are the ultimate source of their actions, unmoved movers of an ontologically distinct kind.

Is it plausible to suppose that we could experience ourselves in agent causal terms? Note that the question is not whether our normal, everyday experience of ourselves as agents should be understood in agent causal terms, rather, the question is whether agent causation is something that could be experienced.

There are two issues that should give us pause before answering this question in the affirmative. First, it is not clear whether the very notion of agent causation is coherent. Searle claims that "It is a constraint on the notion of causation that wherever some object x is cited as a cause, there must be some feature or property of x or some event involving x that functions causally. It makes no sense to say, *tout court*, that object x caused such and such an event." (Searle 2001, 82). Searle's comment concerns the concept of causation, but there is some reason to think that a similar constraint holds for experiences of causation. We are by no means convinced that Searle's objection is sound; we advance it simply as a claim that any development of an agent causal account of the phenomenology of agency must address. Second, even if it is possible to experience oneself as a mover, it does not follow that it is possible to experience oneself as an *unmoved* mover. It is hard to see how one could experience an agent causal relation as undetermined by ones prior states.

We turn now from the question of how the experience of authorship might be understood, to the question of whether such experiences are reliable. Is our experience of ourselves as agents veridical, or is it systematically misleading?

Wegner, among others, has recently argued that there is reason to think that it is misleading (Wegner and Wheatley 1999, Wegner 2002; see also Halligan and Oakley 2000). Wegner's argument rests on the claim that it is possible to create illusions of agency: we can experience ourselves as doing things that we are *not* doing, and we can experience ourselves as not doing things that we *are* doing. Wegner takes such dissociations to show that experiences of agency "may only map rather weakly, or perhaps not at all, onto the actual causal relationship between the person's cognition and action" (Wegner and Wheatley, 1999: 481).

Consider one of Wegner's examples of a dissociation between agency and the experience thereof: his *I-spy* experiment (Wegner and Wheatley 1999). In this

experiment participants and an experimental confederate had joint control of a computer mouse that could be moved over any one of a number of images on a board (e.g., a swan). On certain trials the confederate forced the pointer to land on a target image while participants were primed with the name of the image via headphones at a certain temporal interval either after or before the pointer landed on the image. When the prime occurred immediately before the pointer landed on the target image participants showed an increased tendency to self-attribute the action, that is, to claim that they had intended to land on the image. Wegner and Wheatley argue that the prime creates an experience of agency in the absence of an exercise of agency.

The *I-Spy* experiment does suggest that priming can modulate experiences of agency, but it does not, we suggest, provide much support for *general* skepticism about the reliability of our experience of agency. Indeed, the lengths to which one has to go in order to create a non-veridical experience of agency demonstrates just how reliable those mechanisms that generate the experience of agency are. That they are *fallible* should not be surprising, for all monitoring mechanisms are fallible. Think of visual perception! People do, of course, confabulate their intentions (see Wilson 2002), but there is no reason to think that the confabulation of intentions is more common than perceptual error. Indeed, Wegner's own account of the experience of mental causation suggests that such experiences are generally reliable, for his model predicts that one is likely to experience oneself as acting on the basis of a certain intention only when one is acting on the intention in question (see, further, Bayne forthcoming).

Wegner also offers the indirectness of experiences of agency as evidence against their veridicality. But the mere fact that the relationship between an experiential state and the state of the world that it is monitoring is theoretically mediated gives us no reason to regard the resulting experiences as unreliable. Visual phenomenology is also theoretically mediated, for it relies on assumptions about the structure of the perceptual environment, but this is no reason to assume that visual experience misrepresents the world.

Another possibility is that Wegner is conceiving of the experience of authorship in agent causal terms. Perhaps he is assuming that we experience ourselves as

Cartesian selves - agents who lie outside the causal nexus of the physical world. But agent causal accounts of authorship are not committed to endorsing a Cartesian conception of the self. Rather than conceive of the agent causal relation as holding between a Cartesian self and its actions, it is open for an agent causationist to think of the agent causal relation as holding between an organism and its actions (see e.g. Bishop 1983). That is, one can think of the acting self as identical to the animal. Arguments against Cartesian conceptions of the self would have no impact on this version of agent causation.

A second response to Wegner's position (as here understood) is to allow that experiences of agent causation are non-veridical, but to insist that there are other components of the experience of authorship that are unscathed by the attack on agent causation. As we have seen, it is an open question whether the experience of agent causation is a component of the experience of authorship, and even if it is, it is doubtful that it is the sole component of the experience of agency. We suggest that much of the content of our normal, everyday experience of authorship would survive the discovery that there is no primitive causal relation between the self and its actions.

A further possibility is that Wegner is building certain claims about consciousness or deliberation into what it is for an action to be authored. Certain passages in his writings suggest that Wegner holds that our experiences of authorship are illusory because our actions are generated by unconscious and non-deliberative processes rather than conscious, deliberative processes. On this view, the only actions that are truly our own would be those that originate from processes of conscious deliberation.

We think that this account of the content of the experience of authorship should be rejected. For one thing, it is highly restrictive. Few of our actions derive from processes of conscious deliberation, and there is no reason to think that those actions that are non-deliberative are any less authored than those that are. Of course, it may be that conscious deliberation brings in additional kinds or forms of authorship – it is certainly true that self-consciousness creates new levels and layers of control within an organism – but there is no reason to think that the

content of the quotidian experience of authorship refers to authorship by an essentially conscious deliberator.

But rejecting Wegner's attack on the experience of authorship is far from settling all of the issues raised by such experiences. One question that deserves some attention is how common such experiences are. Do they occur only in some contexts—such as those involving deliberation, decision-making or self-control—or are they a component of all experiences of agency, even those involving stimulus-driven (or automatic) actions? We are inclined towards the latter view. Although it seems to be true that experiences of authorship are recessive or dampened in the context of automatic actions (or, at least, actions which we experience as automatic), we doubt that the phenomenology of authorship is entirely lacking from such experiences (see also Haggard, this volume). There seems to be a stark contrast between the phenomenology of stimulus-driven actions, such as accelerating in response to a green light, and the phenomenology of pathologies of agency, such as the anarchic hand syndrome and Penfield actions (actions produced by direct neural stimulation, by the American neurosurgeon Wilder Penfield). It is tempting to think that this difference is at least in part a difference in the experience of authorship: in the former case one has an experience of authorship (albeit recessive), while in the latter cases one has no experience of authorship. Indeed, it is not unlikely that the experience of authorship is essential to the experience of agency – that to experience a movement as one's own action necessarily involves an experience of oneself as the author of the movement.

We turn now to consider anarchic hand (Della Sala et al, 1991; Goldberg & Bloom, 1990) and Penfield actions in more detail (Penfield and Welch 1951; Penfield 1975), for both syndromes raise acute questions about the content of the experience of authorship. The phenomenology of the anarchic hand appears to be one of alienation authorship – the anarchic hand patient fails to experience their actions as their own. The patient will often describe the anarchic hand as having a will of its own, and this description appears to reflect the phenomenology of the syndrome. It is intuitively plausible to suppose that the experience of alienated authorship in the anarchic hand is veridical: the anarchic

hand patient is *not* the author of ‘their’ anarchic actions (see Peacocke 2003; Levy and Bayne 2004). Penfield’s patients produced a number of actions, such as movements of the hand, or vocalizations – but they did not experience a sense of authorship of the actions; in fact, they experienced their actions as alien and unowned. And again, there is some temptation to judge that this experience of alienated agency is veridical: Penfield patients are not the agents of their actions. But although intuitively plausible, the thought that the phenomenology of alienated authorship in these cases is veridical can be challenged by considering another pair of disorders of volition: utilization behavior (Lhermitte 1983; Estlinger et al, 1991) and Delgado actions (Delgado 1969). Patients with utilization behavior engage in stimulus driven behavior which looks rather similar to that seen in the anarchic hand. Both syndromes involve an inability to inhibit pre-potent movements. (In the anarchic hand the movements are both endogenously and exogenously triggered, while in utilization behavior the triggers are solely exogenous.) Yet despite these similarities, the two syndromes appear to differ phenomenologically: whereas anarchic hand patients have an experience of alienated authorship, patients with utilization behavior appear to experience their utilization actions as their own.⁵ Now, the similarities in the generation of behavior in these two syndromes puts pressure on what we should say about the phenomenology of the two conditions. There is a prima-facie case for thinking that if the phenomenology of the anarchic hand is veridical then that of utilization actions is not. If, on the other hand, we insist that the experience of authorship in utilization behavior is veridical, then there is some reason to think that we should judge experiences of non-authorship in the anarchic hand as non-veridical.

Consider now Delgado actions. Delgado (1969) also elicited “actions” by direct neural stimulation, but unlike Penfield’s patients, Delgado’s patients seem to have experienced the resultant actions as their own. Perhaps it is the phenomenology of Delgado’s patients, rather than Penfield’s patients, which is veridical. Again, there is a prima facie case for thinking that if the phenomenology of authorship in Delgado actions is veridical then the

phenomenology of non-authorship in Penfield actions is non-veridical, and vice-versa.

We take no stand here on which of these experience of agency might be veridical. It is difficult to make a judgment about these cases without knowing more about the disorders and their aetiology. It could, of course, turn out that there are relevant differences between (say) Delgado and Penfield actions, such that the respective experiences of agency and non-agency are both veridical. We present these pairs of cases to draw attention to some of the challenges one faces in giving an account of the experience of authorship, and the need to consider a full range of cases in addressing those challenges.

5. The phenomenology of effort

A third component of the experience of agency is the sense of effort. The world's resistance to our actions, coupled with our limited success in changing it, may give rise to the feeling of effort – the experience of needing to invest energy and will-power in our actions. This experience is perhaps most pronounced in the context of physical agency, but it also occurs in the context of mental agency. We shall focus largely on the effort associated with mental acts.

The experience of effort raises a number of questions. How prevalent is it? Is it a feature of human mental life generally, or is it restricted to a relatively few specific circumstances? What is the relationship between the experience of mental effort and the experience of physical effort? Does the experience of effort have representational content? If so what is the nature of that content? We examine these questions in turn.

Some theorists suggest that the experience of mental effort is restricted to situations of motivational conflict (Holton 2003). On this view, we exert mental effort only when we attempt to resist the pull of our desires. This might be described as the experience of exerting *will-power*, which can be identified with the capacity for reflective choice against the momentum of the impulsive system (Bechara, this volume). We suggest, however, that the experience of mental effort extends beyond cases of motivational conflict. Mental effort is also experienced when we actively direct our thoughts. Anyone who has struggled with a difficult

conceptual issue has experienced the effort involved in thinking a problem through. It gives rise to characteristic feelings of tiredness and a growing urge to stop. When we do stop for a break, it seems to require real effort to return to the task.

Of course, we might try to assimilate this kind of case to the experience of motivational conflict. As we grow tired, we experience both a desire to rest as well as a desire to continue to work on the puzzle, and it is the conflict between these two desires that leads to the experience of mental effort. However, it seems better to understand the mental effort involved here as distinct from the effort involved in resisting a desire. Effort seems to be involved not only in motivating the urge to stop, but also in *causing* it: tasks which involve concentration, for instance, are more tiring than tasks which don't, other things equal, and it is plausible to suggest that this is because they are more effortful.

Clear and vivid experience of mental effort is perhaps restricted to cases of motivational conflict and active direction of thought. However, there is some reason to think that a recessive experience of effort accompanies a very wide range of mental actions: Hard decisions are experienced as requiring effort, perhaps as a consequence of the cognitive resources we need to devote to them. There is also some plausibility to the idea that even easy decisions might require *some* degree of effort (Mele 1987).⁶ Consider in this respect the following question: do we experience a sense of effort only for physically demanding actions, or do we experience a sense of effort in proportion to the demands of the action (and the tiredness of our muscles)? There seems to be plausible grounds for asserting the latter: someone feels a great sense of effort when they lift 60 kilograms, a lesser but still clearly perceptible sense of effort when they lift 15 kilograms, and less still when they lift 5 kilograms. Perhaps it is generally true that the sense of effort dwindles proportionally as the resources – be they mental or physical – needed for the task decrease. However, it does seem to be true that there are some actions that are experienced as effortless (Marcel 2003).

We turn now to the relationship between the experience of physical and mental effort. In both cases, the experience of effort seems to be the experience of resisting a force: usually an external force in the physical case, internal in the

mental case. Despite the apparent difference in the nature of the force being resisted, it may seem that the experience of resisting it is very similar.

Some theorists argue that physical and mental forces cause human movements in very different ways. They point out that a physical force can act directly on my body—the wind can lift me off my feet, for instance—but mental forces cannot. Instead, mental forces operate on our minds by *motivating* actions rather than *compelling* them. And since physical forces compel but mental forces do not, mental forces can be resisted indefinitely. As Feinberg puts it, “Human endurance puts a severe limit on how long one can stay afloat in an ocean, but there is no comparable limit to our ability to resist temptation” (Feinberg 1970, 283). It may be that this alleged difference gives rise to differences in the phenomenology of resisting a mental as opposed to a physical force.

Consider the experience of giving in to a mental force, when, for instance, we find ourselves breaking a resolution in the face of temptation. In these cases, we don’t experience ourselves as compelled by mental forces. Instead, we seem to have an experience of authorship. As Holton remarks, “It certainly doesn’t feel as though in employing will-power one is simply letting whichever is the stronger of one’s desires or intentions have its way. It rather feels as though one is actively doing something, something that requires effort.” (2003: 49)

Watson (2004) suggests that the experience of resisting mental forces in cases of motivational conflict is not an experience of *compulsion*—as in the case of physical effort—but an experience of *seduction*. Whereas a physical force can *bypass* someone’s will, overcoming their muscular strength for example, desires work *through* our will. In cases of motivational conflict we are unable to make a wholehearted effort to resist mental forces. We do not try as hard as we might to resist the desires in question because they are our own; we are divided against ourselves. Watson claims that this difference between mental and physical effort is reflected in the phenomenology of giving in to temptation: “one who is defeated by appetite is more like a collaborationist than an unsuccessful freedom fighter. This explains why it can feel especially shameful” (2004, 65-6).

There are at least two problems with Watson's suggestion. First, although seduction (and therefore reduced effort) might be characteristic of motivational conflict, it does not seem to characterise all instances of mental effort. Watson's account does best in contexts in which one struggles to overcome a desire that one clearly identifies as one's own. It works less well in contexts in which agents resist desires that they experience as alien to themselves, such as certain cases of Tourette's Syndrome and Obsessive-Compulsive Disorder. In such cases agents can experience intense urges to engage in self-destructive actions—such as punching themselves—that they take themselves to have no reason at all to perform. It seems unlikely that there is *any* element of seduction in these cases. Although the force being resisted is (in some sense) internal to the agent, it is plausible to suppose that it is experienced as no less alien or external to the agent than physical forces.

Second, we doubt that the phenomenon of seduction is restricted to contexts involving mental effort. Arguably, many cases of being overwhelmed by physical forces also involve some degree of seduction. As a physical effort becomes increasingly difficult, we experience a temptation to stop. As a result, we are likely to experience motivational conflict when engaged in physical effort; if motivational conflict prevents us from making a wholehearted effort, as Watson suggests, then this is likely to be a regular feature of resistance to physical forces.

More generally, the experience of being overcome by a physical force is typically quite unlike the experience of being lifted off one's feet by the wind. Consider the experience of muscular fatigue – for instance, the experience of jogging until you are too exhausted to continue. We suggest that most of us experience succumbing to physical exhaustion as a matter of voluntarily giving to external forces – that is, we do not experience ourselves as literally unable to continue at the task.

Sometimes, however, people push themselves beyond the point at which giving in feels voluntary. Professional athletes, for instance, sometimes exert themselves to the point where their legs will no longer support their weight. This experience is very different from the experience of resisting a mental force, but it may well

resemble the experience of mental collapse, when we feel as though our minds are no longer under our control. (If legs can collapse, minds can become delusional). Both the normal and the extreme cases seem to support a general parallelism between the experiences of mental and physical effort.

We turn now to the representational content of the experience of effort. One might be tempted to think that the experience of effort has no representational content at all. Perhaps it is a mere feeling – a brute sensation that has no veridicality conditions. Although we are certainly happy to grant that experiences of effort have a phenomenal character that outruns their representational content, we think that it would be a mistake to deny that the experience of effort has *any* representational content. In exerting effort it feels as though one is exerting a power of one's own against a force. (Whether or not this force feels alien or merely unwanted differs from case to case, with motivational conflict tending toward the unwanted end of the spectrum and pathologies such as Tourette's Syndrome and OCD tending towards the alien end). Moreover, it feels as though the power one is drawing on becomes progressively weaker as effort is exerted. In the physical case, we know (or think we know) the nature of that power: it is muscular strength. We suggest that one is drawing on an analogous power in exercising mental effort, and that the experience of mental effort involves a representation of the utilization and progressive fatigue of *mental muscles* – what Baumeister and colleagues call *will-power* (Baumeister, Bratslavsky et al. 1998; Muraven, Tice et al. 1998; Baumeister 2002). Moreover, we suggest that our experiences of effort are broadly consistent with what Baumeister et al. have discovered about will-power.

First, will-power is depletable over the short-term. Subjects required to perform a task which requires the application of will-power perform worse at subsequent tasks requiring will-power. Moreover, the effect seems independent of (physical) fatigue: controls who are required to perform tiring tasks which do not require will-power do not suffer the same degree of impairment at subsequent will-power tasks.

We have also suggested that mental effort is experienced in cases of motivational conflict, as well as in decision-making more generally, and that the experience is

of utilizing a mental muscle. Ego-depletion studies seem to support all three claims. (1) They provide evidence that there is a depletable resource – akin to a mental muscle – used in resisting temptation. This evidence takes the following forms:

Behavioral evidence: Ego-depleted individuals quit tasks requiring persistence earlier than do non-depleted individuals. Of course, there may be alternative explanations of why they quit, but it seems plausible to suggest that the effort of the task is an important factor.

Physiological evidence: Subjects engaged in ego-depletion tasks exhibit the same kinds of physiological arousal which accompany physical effort, such as a rise in pulse rate, blood pressure, and skin conductance (Muraven, Tice et al. 1998).

Self-reports: Though subjects in ego-depletion experiments report no more fatigue than controls after the initial ego-depletion task, they report greater levels of fatigue after performing the common task (at which they persist for a shorter time than controls) (Vohs, Baumeister et al. Unpublished). This is consistent with their having exerted greater effort than controls.

This evidence seems to suggest that persistence in some tasks does indeed take effort, and that the ability to exert this effort diminishes with continuous use. This goes some way to showing that the experience of mental effort as the experience of progressive exhaustion of a depletable resource is veridical.

(2) Ego-depletion studies also seem to indicate that the experience of effort in decision-making is veridical, inasmuch as it depletes the same resource used in resisting temptation. Close-call decisions seem to be ego-depleting (Vohs, Baumeister et al. Unpublished). Further evidence for this claim comes from the self-reports of subjects in introspectionist studies of decision-making (Nahmias et al. 2004). At least when decisions are not routine, there seems to be an experience of effort, and this experience may accurately represent the fact that these mechanisms are more energy expensive than those which implement automatic decision-making.⁷ (3) As already noted, the ego-depletion hypothesis seems to

suggest that the resources utilized are progressively weakened by continuous use, in a manner akin to the way muscular strength is depleted.

But although ego-depletion experiments go some way toward establishing the veridicality of experiences of mental effort, other studies seem to suggest that some experiences of fatigue are non-veridical. (The research we draw on here involves the perception of muscular fatigue, but it is entirely possible that similar results can be obtained for experiences of mental fatigue.) Research by St Clair Gibbon and colleagues seems to indicate that the degree of fatigue we experience is “teleoanticipative” – that is, it is influenced by unconscious representations of the amount of muscular effort that will be required in the near future, rather than merely being a response to the depletion of a physical resource. Whereas traditional theories explain muscular exhaustion in terms of the total depletion of a physical resource, the teleoanticipation view suggests that exhaustion is the product of neural mechanisms designed to preserve physical resources in case of emergencies (Hampson, St Clair Gibbon et al. 2001; Noakes, St Clair Gibbon et al. 2004).

The teleoanticipation explanation of the experience of fatigue suggests that its content is non-veridical, if, as seems plausible, the experience of fatigue represents ourselves as having entirely “run out of gas.” On this view the mechanisms causing the perception of fatigue are designed to preserve resources well above the critical level, and it seems natural to think of them as generating systematically misleading representations of our energy levels.

But perhaps the experience of exhaustion does not represent the complete depletion of resources needed for mental and physical effort. Perhaps, instead, the experience represents an inability to make the effort in the precise circumstances in which we believe ourselves to be. And, on the teleoanticipation view, this experience may well be veridical, for the view suggests that the resources we have when exhausted can be accessed only when needed.

The research supporting the teleoanticipation view of fatigue has been restricted to the experience of muscular fatigue. However, we suggest that the demonstration that physical fatigue is not the veridical perception of a

catastrophic muscular event provides some indirect support for our claim that the experiences of mental and physical effort and fatigue are closely analogous. The temptation to believe that there is a deep gulf between the two kinds of experiences seems to be the product of the belief that although physical forces are subject to total depletion, mental forces are not. The demonstration that physical fatigue is not normally a product of running too low on physical resources goes some way towards making the view that mental fatigue is analogous to physical fatigue more palatable. Though it is probably true that people typically give in to mental forces while still possessing enough mental resources to hold out longer, this is normally the case with regard to physical effort as well. In both cases, agents do not experience themselves as able to draw upon (or perhaps as possessing) resources which are nevertheless present.⁸

We speculate that ego-depletion triggers the same kind of mechanisms that give rise to the perception of physical fatigue, prompting the subject to desist from a task in order to save energy for emergencies. Perhaps in mental effort a mediating role is played by the inhibition mechanism that prevents internal and external stimuli from leading to action (the frontal lobe mechanisms damaged in both utilization behavior and anarchic hand syndrome). We suggest that operating this inhibiting mechanism takes energy and is therefore experienced as effortful. Evidence that it is the energy utilized by the inhibitory mechanism that is depleted comes from the behavior of ego-depleted subjects. They do not lapse into passivity when they are ego-depleted—as we would expect if ego-depletion used up a general-purpose energy supply—but instead act on their strongest desires, whether these are for rest or for something else. For instance, ego-depleted dieters eat more than do ego-depleted non-dieters (Vohs and Heatherton 2000).

A final question: How does the experience of effort relate to the experience of authorship and mental causation? Although the experience of authorship is, we have suggested, a central component of the phenomenology of agency generally, it appears to be particularly vivid in experiences of effort. The experience of effort involves an experience of the self as a source of force. Arguably, if the experience of agency were limited to experiences of mental causation, we would

not experience ourselves as playing an active role in endorsing or resisting our urges. We might experience conflict between competing desires and intentions, but we would not experience ourselves as allied with some desires and opposed to others. If the experience of authorship ever takes the experience of agent causation we suggest that it is in contexts in which the experience of effort is particularly vivid.

6. The Phenomenology of Agency and Disorders of Volition

In this final section we return to the connection between the phenomenology of agency and disorders of volition. There are two perspectives one can adopt on this connection. The first perspective considers the ways in which disorders of volition involve departures from the normal phenomenology of agency. Such departures can involve the absence of content that is usually present in the phenomenology of agency, and they can involve the presence of content that is usually absent from the phenomenology of agency. The experience of being caused to act by one's desires – which might occur in some cases of addiction and OCD – would be an example of the latter form of abnormality, and the lack of the experience of authorship in the anarchic and Penfield cases would be an example of the former. Of course, not all unusual experiences of agency ought to count as disorders in the phenomenology of agency. Flow experiences, which seem to involve a departure from the normal phenomenology of agency, do not seem to count as disorders of volition.

A second perspective on the relationship between the phenomenology of agency and disorders of volition is in terms of the operation of those mechanisms responsible for generating the phenomenology of agency. Here it is useful to distinguish two ways in which unusual experiences of agency can arise. On the one hand, abnormal experiences of agency might be a reflection of the fact that the agent's actions are themselves abnormal in etiology and structure. If a person's action generation system is abnormal and their phenomenology of agency generating system is normal then we would expect them to have abnormal experiences of agency; – at least, so long as the system that generates the phenomenology of agency is designed to track the structure of the agent's actions. On the other hand, the etiology and structure of the agent's actions

might be perfectly normal, and the abnormality in the agent's experiences of phenomenology might be due solely to the malfunctioning of those mechanisms responsible for generating such experiences. (A third possibility is that the mechanisms responsible for generating actions and those responsible for generating the experience of agency are both malfunctioning.) Of course, this tidy distinction between pathologies of action-generation and pathologies in the generation of the phenomenology of agency is complicated by the fact that there are undoubtedly intimate feedback loops between the phenomenology of agency and the generation of actions: abnormalities in the experience of agency are likely to change the structure of the action itself.

Consider this distinction as it applies to hypnotic actions. It could turn out that the mechanisms governing hypnotically-suggested actions are functioning entirely normally, and that the only pathology here is in the patient's *experience* of their movements (Kirsch and Lynn 1997; Haggard et al., 2004). Alternatively, it might be the case that the mechanisms that are disrupted are those governing the production of the hypnotically-suggested actions themselves, and that the patient's experiences of involuntariness are an accurate reflection of the fact that the actions in question are not produced by those mechanisms governing the voluntary production of behavior (Woody & Bowers 1994). And, of course, a third possibility is that the mechanisms responsible for the generation of the movements and the phenomenology of agency are both malfunctioning.

There is much about the phenomenology of agency that is obscure. Perhaps this reflects the relative neglect of the topic; perhaps it arises from the fact that the phenomenology of agency appears to be less vivid and stable than the phenomenology of perception. In this paper we have attempted to fill in some of the structure of experience of agency. As we have seen, much remains to be done before we know, in the relevant sense, what it is like to be an agent.

References

- Bach, K. 1978. A Representational Theory of Action. *Philosophical Studies* **34**: 361-379.
- Baumeister, R. F. 2002. Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function. *Self and Identity* **1**: 129-136.
- 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.
- Bayne, T. forthcoming. Phenomenology and the Feeling of Doing: Wegner on the Conscious Will. *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*. S. Pockett, W. P. Banks and S. Gallagher. (eds.) Cambridge, MA: MIT Press.
- Bechara, A. Broken Willpower: Impaired Mechanisms of Decision-Making and Impulse Control in Substance Abusers
- Bishop, J. 1983. Agent Causation. *Mind*, 92: 61-79/
- Bliss, J. 1980. Sensory experience of Gilles de la Tourette syndrome. *Archives of General Psychiatry*, 37: 1343-7.
- Chisholm, R. 1982. Human Freedom and the Self. *Free Will*. G. Watson. (ed) Oxford: Oxford University Press: 24-35.
- Cohen, A. and Leckman, J.F. 1992. Sensory Phenomena Associated with Gilles de la Tourette's Syndrome. *Journal of Clinical Psychiatry* **53**: 319-23.
- Davidson, D. 1973. Freedom to Act. *Essays on Freedom of Action*. T. Honderich. (ed) London: Routledge: 137-56.
- Delgado, J.M.R. 1969. *Physical control of the mind: Toward a psychocivilized society*. New York: Harper and Row.
- Della Sala et al, S., Marchetti, C., Spinnler, H. 1991. Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia* 29(11): 1113-1127.
- Estlinger, P.J., Warner, G.C., Grattan, L.M., Easton, J.D. 1991. Frontal lobe utilization behavior associated with paramedian thalamic infarction, *Neurology*, 41: 450-52.
- Feinberg, J. 1970. What is so Special About Mental Illness? *Doing and Deserving*. Princeton, Princeton University Press: 272-92.
- Frith, C. 2002. Attention to action and awareness of other minds. *Consciousness and Cognition* **11**: 481-87.

- Ginet, C. 1990. *On Action*. Cambridge, Cambridge University Press.
- Goldberg, G. & Bloom, KK 1990. The alien hand sign. Localization, lateralization and recovery. *American Journal of Physical Medicine and Rehabilitation*, 69: 228-38.
- Graham, G. Forthcoming. Self-seeking in action. *Consciousness and Cognition*.
- Haggard, P., Cartledge, P., Dafydd, M., Oakley, D.A. 2004. Anomalous Control: When 'free-will' is not conscious. *Consciousness and Cognition* 13: 646-654.
- Haggard, P. This volume. Conscious intention and the sense of agency.
- Halligan, P. and Oakley, D.A. 2000. Greatest Myth of All. *New Scientist* 168 (2265): 35-39.
- Hampson, D.B., St Clair Gibbon, A., Lambert, E.V., and Noakes, T.D. 2001. The Influence of Sensory Cues on the Perception of Exertion During Exercise and Central Regulation of Exercise Performance. *Sports Medicine* 31: 935-952.
- Hécaen, H., Talairach, J., David, M., Dell, M.B. 1949. Coagulations limitées du thalamus dans les algies du syndrome thalamique: résultats thérapeutiques et physiologiques, *Revue de Neurologie*, 81: 917-31.
- Holton, R. 2003. How is Strength of Will Possible? *Weakness of Will and Practical Irrationality*. S. Stroud and C. Tappolet (eds.) Oxford, Clarendon Press: 39-67.
- Horgan, T., Tienson, J., & Graham, G. 2003. The Phenomenology of First-Person Agency, *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. S. Walter and H-D Heckmann (eds.) Exeter, UK: Imprint Academic: 323-40.
- Kirsch, I. and S. J. Lynn 1997. Hypnotic involuntariness and automaticity of everyday life. *American Journal of Clinical Hypnosis* 40: 329-48.
- Levy, N. and Bayne, T. 2004. A Will of One's Own: Consciousness, Control and Character. *International Journal of Law and Psychiatry* 27: 459-470.
- Lhermitte, F. 1983. Utilization behavior and its relation to lesions of the frontal lobes. *Brain*, 106: 237-55.
- Libet, B. 2004. *Mind Time*. Cambridge, Mass., Harvard University Press.
- Marcel, A. 2003. The Sense of Agency: Awareness and Ownership of Action. *Agency and Self-Awareness: Issues in Philosophy and Psychology*. J. Roessler and N. Eilan (eds). Oxford, Clarendon Press: 48-93.
- Mele, A. 1987. *Irrationality: An Essay on Akrasia, Self-Deception and Self-Control*. New York: Oxford University Press.

- Muraven, M., D.M. Tice, et al. 1998. Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology* **74**: 774-789.
- Nahmias, E., S. Morris, et al. 2004. The Phenomenology of Free Will. *Journal of Consciousness Studies* **11**(7-8): 162-79.
- Noakes, T. D., A. St Clair Gibbon, et al. 2004. From Catastrophe to Complexity: a Novel Model of Integrative Central Neural Regulation of effort and fatigue during exercise in humans. *British Journal of Sports Medicine* **38**: 511-514.
- O'Connor, T. 1995. Agent Causation. *Agents, Causes, Events*. T. O'Connor (ed). New York: OUP: 173-200.
- Pacherie, E. 2000. Intentions in action. *Mind and Language*, **15**: 400-32.
- Paus, T., Koski, L. Caramanos, Z., and Westbury, C. 1998. Regional difference in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *Neuroreport*, **9**: R37-R47.
- Peacocke, C. 2003. Action: Awareness, Ownership, and Knowledge. In J. Roessler and N. Eilan (eds.) *Agency and Self-Awareness*. Oxford: OUP: 94-110.
- Penfield, W. 1975. *The Mystery of Mind*. Princeton: Princeton University Press.
- Penfield, W. and Welch, K. 1951. The Supplementary Motor area of the cerebral cortex. *Archives of Neurology and Psychiatry* **66**: 289-317.
- Searle, J. 2001. *Rationality in Action*. Cambridge, Mass., Harvard University Press.
- Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge, Cambridge University Press.
- Taylor, R. 1966. *Action and Purpose*. Englewood Cliffs, N.J., Prentice-Hall.
- Vohs, K. D., R. F. Baumeister, et al. Unpublished. Self-Regulation and Choice.
- Vohs, K. D. and T. F. Heatherton (2000). Self-Regulatory Failure: A Resource-Depletion Approach. *Psychological Science* **11**: 249-254.
- Wakefield, J. and Dreyfus, H. 1991. Intentionality and the Phenomenology of Action. In E. Lepore and R. van Gulick (eds) *John Searle and His Critics*. Oxford: Blackwell (259-70).
- Watson 2004. Disordered Appetites: Addiction, Compulsion, and Dependence. In *Agency and Answerability: Selected Essays*. Oxford: Clarendon Press. 59-87.
- 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 Psychologist*, 54: 480-91.
- Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.
- Woody, E.Z., & Bowers, K.S. 1994. A frontal assault on dissociated control. In S. J. Lynn & J.W. Rhue (Eds.) *Dissociation: Clinical and Theoretical Perspectives* (pp. 52-79). New York: Guilford.
- Zhu, J. 2004. Locating Volition. *Consciousness and Cognition* 13: 302-22.

¹ The authors gratefully acknowledge David Chalmers, George Graham, Peter Menzies, Elisabeth Pacherie, Kathleen Vohs and audiences at the Department of Philosophy, Melbourne University and the Department of Philosophy (RSSS), The Australian National University, for helpful comments on this paper and the issues discussed therein. The research was partially funded by an Australia Research Council Discovery Grant, DP0452631.

Notes

² Other putative experiences involved in the phenomenology of agency include: the experience of *freedom*; the experience of *trying*; and the experience of *deliberation* or *decision-making*. Although these experiences are closely related to the experiential states we discuss, we lack the space to address the connections here.

³ A related objection derives from Hécaen et al.'s data (Hécaen et al., 1949; reported in Marcel 2003). Hécaen and colleagues stimulated the central thalamus, which produced contralateral hand clenching and unclenching. Hécaen's patients seem to have experienced these movements as their actions, despite having no idea why they had made them. These data suggest that one can experience a movement as an action without experiencing it as implementing an intention.

⁴ It is often thought that such disorders as Tourette's Syndrome involve an experience of being caused to act by one's urges, but individuals with Tourette's report experiencing their tics as having their source in themselves (Bliss 1980; Cohen and Leckman, 1992).

⁵ The phenomenological differences are nicely illustrated by a case in which a patient exhibited utilization behavior with his right hand and anarchic agency with his left hand: the patient was unconcerned about the former but troubled by the latter! (Marcel, 2003)

⁶ Note that it is important not to confuse a weak (recessive, peripheral) experience of effort with an experience of weak effort. The content of an experience is one matter, its phenomenal saliency is another.

⁷ Many studies have shown that the prefrontal cortex and the anterior cingulate cortex are highly active in close-call decision-making, but not in more automatic tasks (Paus, Koski, et al. 1998). As subjects become proficient at a task, the degree of activation of the ACC, in particular, decreases. See Zhu (2004) for a review. The experience of effort may be a reflection of the extent to which certain tasks are energy-intensive.

⁸ We note one further implication of the claim that the sensation of effort, including the sense that exhaustion is overwhelming, is a product of teleoanticipation: the popular method of ascertaining ability or capacity by a straightforward appeal to counterfactuals will give the wrong result in many cases. We cannot infer from the fact that subjects would find the reserves to persist in a task in some counterfactual circumstances that they can persist in the actual circumstances.