Philosophy Compass 5/9 (2010): 793–805, 10.1111/j.1747-9991.2010.00346.x

Philosophical Questions about the Nature of Willpower

Chandra Sekhar Sripada*
University of Michigan

Abstract

In this article, I survey four key questions about willpower: How is willpower possible? Why does willpower fail? How does willpower relate to other self-regulatory processes? and What are the connections between willpower and weakness of will? Empirical research into willpower is growing rapidly and yielding some fascinating new findings. This survey emphasizes areas in which empirical progress in understanding willpower helps to advance traditional philosophical debates.

1. Introduction

Consider the dieter who so much wants to lose weight, but when the dessert cart rolls by, succumbs. Or the lush who promises himself that he will have just one glass of wine, but in the end can not resist having a second and a third. Or even the 10th year graduate student who is supposed to be writing her dissertation, but caves in when her friends pressure to her to go to the party. Folk practice explains the failure of these agents to achieve what they had originally set out to do in terms of the absence or insufficiency of *willpower.* But what exactly is willpower and how does it work? Asking this question immediately raises fascinating issues at the intersection of philosophy and a number of adjoining empirical disciplines including psychology, neuroscience, and psychiatry. This article surveys four key questions about the nature of willpower that are particularly interesting from a philosophical perspective. Some of the questions posed in this survey are addressed in the current literature in only formative or tentative ways. Thus, this survey of the literature also serves as an agenda for future research.

Question 1: How is Self-Control Possible? How can the Self Control Itself?

Common sense understands willpower as a form of self-control. An agent exercising willpower attenuates or suppresses one of her own desires, or in some other way prevents that desire from winning control over action. Willpower is a form of *synchronic* self-control. In this form of self-control, an *occurrent* desire is prevented from prevailing in action. Synchronic self-control can be contrasted with diachronic control, in which an agent seeks to prevent some non-occurrent future desire from driving action. For example, suppose right now Charlie most desires to stay on his diet and he does not have a desire to eat a slice of cake. But he knows that were he to go to the pastry shop with his friends, the sight of the cherry chocolate cake would trigger a desire for the cake to which he is sure he will succumb. So Charlie decides to stay at home rather than go to the pastry shop with his friends, and in this way ensures that the desire to eat the cherry chocolate cake will not be triggered. But in preventing this potential future desire from becoming active and prevailing in action, Charlie does not exercise willpower.

Willpower, as a species of synchronic self-control, requires that the desire that is rendered motivationally inefficacious be active *at the very time* willpower is exercised.

Some philosophers have found the idea of synchronic self-control (and *a fortiori*, exercises of willpower) puzzling. The starting place for this puzzle is the observation that there is a tight connection between what an agent most desires to do and what an agent will do. We can put this idea in a slightly more rigorous form:

> (1) If an agent most desires to perform some action *x*, and if she believes herself free to *x*, then she will *x*, if she does anything at all intentionally.

If we grant (1), then this leads to a dilemma about willpower[1] that we can illustrate by means of another example. Suppose Charlie is on a diet, but one evening he goes to the pastry shop with his friends. Now he finds his strongest desire is to eat an enormous slice of cherry chocolate cake. If eating the cake is indeed his strongest desire, then how can Charlie exercise willpower to try to not eat the cake? Given (1), and given that exercising willpower is not what Charlie most desires, then Charlie will eat the cake and will not exert willpower to try to not eat the cake. But suppose then, and this takes us to the other horn of the dilemma, that eating the cake is *not* Charlie's strongest desire. That is, Charlie has a strong desire to eat the cake, but the desire to stay on his diet remains his *strongest* desire. In this case, there doesn't seem to be any need for willpower. Given (1), and given that staying on the diet remains Charlie's strongest desire, then Charlie will stay on his diet. There is no special need to invoke the exercise of willpower to explain why Charlie stays on his diet and does not eat the cake.

One solution to this 'puzzle of synchronic self-control' is to deny that exercises of willpower are properly called actions. According to Jeannette Kennett and Michael Smith (Kennett and Smith, 'Frog and Toad Lose Control'; Kennett and Smith, 'Synchronic Self-Control Is Always Non-Actional'), willpower consists in an agent's dispositions to have certain thoughts – thoughts that highlight certain considerations while deemphasizing others, thus overall enhancing (or blunting) the strengths of the agent's desires. For example, at the moment that Charlie most desires to eat the chocolate cake, suppose Charlie experienced thoughts that the cake is a large lump of fat accompanied by images of the fat curdling in his stomach. The occurrence of 'desire-modifying thoughts' such as these would weaken Charlie's desire to eat the cake and in this way allow Charlie's desire to stay on his diet to prevail. Now, if Charlie needed to *act* to bring about these desire-modifying thoughts, then the puzzle of synchronic self-control would arise anew and we might wonder how Charlie can be motivated to summon up these desire-modifying thoughts when his strongest desire is to eat the cake. But, according to Kennett and Smith, desire-modifying thoughts are not brought about by means of an agent's actions. Rather, thoughts such as these are *happenings*, and they are disposed to happen in a particular agent at a particular time precisely to the degree that that agent is rational at that time. Because desire-modifying thoughts are not the products of actions that Charlie undertakes, Charlie can experience desire-modifying thoughts that counteract his motivation to eat the cherry chocolate cake, even when the desire to eat the cake is his strongest desire.

A potential problem for Kennett and Smith is that it seems perfectly possible for a person to intentionally call to mind desire-modifying thoughts at the time of temptation, and to do so for the deliberate purpose of attenuating the temptation-directed desire. Indeed, it is the hallmark of certain forms of psychotherapy (Beck; Beck et al.) that a person should *deliberately* challenge thoughts associated with problematic emotions and desires. A second approach to the puzzle of synchronic self-control that perhaps does a better job in making sense of *active* exertions of willpower is based on the idea that the mind is partitioned into

distinct motivational compartments. While philosophers have historically pursued 'divided mind' approaches that are rooted in Platonic and Freudian (Davidson) thinking, an updated version of a divided mind view might be naturally developed using the resources of *dual-process models* in contemporary psychology (Chaiken and Trope 1999; Kahneman),[2] which are now well-accepted as explanations for how information processing in the mind/brain is organized in a host of domains (e.g., Chaiken, Liberman, and Eagly; Sloman; Stanovich and West; Gilovich, Griffin, and Kahneman).

In general form, a dual-process model postulates that information processing in some psychological domain is subserved by two distinct systems. One system (often referred to as 'system 1', the terminology is from Stanovich and West) is relatively fast, automatic, performs relatively simple associative operations, and has access to only limited information. The other system (often referred to as 'system 2') is slow, consciously controlled, uses linguistic/logical representations, and has access to much larger and more global stores of information. An additional feature of dual-process models is that they typically postulate that domain-relevant information is processed by both systems 1 and 2 *simultaneously*. Perhaps most crucial for our purposes, at least some kinds of dual-process models propose that when these two systems diverge in terms of their outputs, system 2 can exert *regulatory control* over system 1 (Gilbert). That is, the two systems are related by an inhibition mechanism activated by system 2 that overrides, suppresses, or modulates outputs from system 1.

Many theorists from diverse parts of the behavioral sciences have developed dual-process models of motivation in which, very roughly, emotions and urges occupy the role of system 1, while planning/practical reasoning systems occupy the role of system 2 (Metcalfe and Mischel; Loewenstein; Bechara; Sanfey et al.; Hofmann, Friese, and Strack). If we accept these theorists' suggestions that a dual-process structure underlies decision-making and motivation, then we have a natural way of addressing the puzzle of synchronic self-control. Suppose that two distinct motivational compartments within Charlie reach divergent motivational verdicts about what Charlie should do. The strongest desire within the 'system 1' compartment is that Charlie should eat the cake. The strongest desire within the 'system 2' compartment is that Charlie should stay on his diet. On this picture, willpower is naturally understood as the regulatory control mechanism by which the system 2 compartment suppresses or overrides the system 1 compartment. The divided mind model circumvents the puzzle of synchronic self-control because there is no single agent who *both* most wants to eat the cake and simultaneously most wants to resist eating the cake. Rather, within the system 2 compartment, the strongest desire is to stay on the diet, and this system initiates and maintains willpower. The desire to eat the cake is strongest only in a distinct motivational compartment, and this second compartment is not the *agent* that initiates and maintains willpower, but rather the *patient* whose motivational force is suppressed by willpower.

The idea that willpower is a mental action initiated by one part of a divided mind appears either explicitly (Baumeister, Heatherton, and Tice; Loewenstein), or implicitly in much of the recent psychological literature on the subject. The divided mind picture is also attractive because it captures many *common sense* features of willpower, such as that in situations like Charlie's the agent feels divided, there is an active inner struggle between parts of the agent, and willpower is an *action* performed for the express purpose of curtailing the wayward desire. But the divided mind view also raises many questions about how agency should be understood given such a picture. For example, if the divided mind view is correct, then exercises of willpower aren't truly performed by the agent as a whole, but rather are undertaken by only part of the agent in which only a

strict subset of the agent's full set of desires are active. Common sense understands actions as typically brought about by *the agent*, not part of the agent, so the existence of 'sub-personal actions' of this sort requires a careful philosophical defense and explication. Another question concerns the notion of an agent's *strongest desire*. While there are different accounts of how to understand this notion even on a picture of the mind that is not divided [see, for example, the discussion in (Mele, *Motivation and Agency* Ch. 7)], an additional set of problems emerge on a divided mind picture. Since on this view, there are two motivational compartments, there are hence two strongest desires (one in each compartment), and it is unclear which of these desires should be regarded as *the agent's* strongest desire. Thus, a divided mind picture, although attractive in some respects, surely raises as many questions as it answers.[3]

Question 2: Why Does Willpower Fail?

It is plausible that some failures of willpower are due to insufficient motivation. Suppose Charlie and Marley both decide to go on a diet. Charlie's desire to stay on the diet is very strong as he is deeply worried about dire health consequences that will occur if he fails to keep to his diet. Marley's desire to stay on the diet is not so strong – he started the diet in a moment of vanity and while he wants to be a bit thinner, he has no other motivation for wanting to maintain the diet. One day both go to the pastry shop and each considers having a slice of his beloved cherry chocolate cake. Charlie puts vigorous effort into resisting the desire to eat the cake, with success – he leaves the shop with his diet intact. Marely puts only perfunctory effort into resisting the desire to eat the cake, and in the end, fails – he eats an embarrassingly large slice of cake. Here, it is plausible that Marley's failure in his attempt at willpower owes to the fact that his desire to stay on the diet is relatively weak, thus making him correspondingly only weakly motivated to exercise willpower in resisting a contrary desire.

The case of Charlie and Marley raises the question of whether in *all* cases in which willpower fails, the explanation can be traced back to deficiencies in the strength of the agent's motivation to exercise willpower. Suppose instead that our story above ends with Charlie too succumbing to temptation and eating a large piece of cake. Would the fact that Charlie succumbs to temptation permit the inference that, *despite appearances*, Charlie is in fact only weakly motivated, or at least insufficiently motivated, to stay on his diet? An inference of this sort might be blocked if at least some failures of willpower are due not to insufficiency in an agent's underlying motivation to exercise willpower, but rather stem from relatively 'fixed ceilings' on the efficacy of the willpower process itself. Richard Holton offers a helpful analogy with *a runner's speed* to help separate these two notions (Holton 132). If you want to know how fast a runner can run a mile, you will need to know the runner's motivational state, since wanting to run faster usually translates into greater running velocity. But this relationship only holds over a certain range. Once a threshold of speed is crossed, additional motivation to run faster will not lead to greater speed. Even if the agent wants to run 100 miles per hour more than anything at all, his speed cannot exceed the ceiling established by the physical condition of his body. Moreover, this ceiling is relatively fixed, in that if it is changeable at all, it is only changeable through specific training processes that typically extend over long periods of time.

A broadly similar distinction may be applicable in the case of willpower. In particular, we must distinguish the motivational 'inputs' into the mechanisms that implement willpower from fixed ceilings in the efficacy of these mechanisms themselves. Motivation certainly provides part of the explanation for whether willpower will succeed or fail.

Marley, who is only weakly motivated to resist his wayward desire is, other things being equal, more likely to fail at resistance than Charlie. But the mechanisms that implement willpower may also exhibit fixed ceilings in their efficacy that obtain independently of the agent's motivation. Much like the runner for whom motivation alone will not permit speeds of 100 miles per hour, motivation alone may not permit an agent to resist a sufficiently strong wayward desire.

Some philosophers and legal theorists have been skeptical of the idea that there are fixed ceilings in the efficacy of willpower that hold, irrespective of the agent's motivation to resist.

> Strictly speaking no impulse is irresistible; for every case of giving in to a desire… it will be true that, if the person tried harder, he would have resisted it successfully… 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 283)

Folk practice sometimes uses analogies with muscles, fatigue, and endurance to characterize how there might be fixed ceilings in the efficacy of willpower. For example, a person exercising willpower to battle a temptation-directed desire might say 'It overpowered me', 'I could not fight it off', or 'I held out for a long time, but in the end it got the best of me'. In the passage above, Feinberg appears to reject this analogy, as he denies that limitations on willpower can be likened to the limits on endurance that are associated with muscular activity. However, recent studies in psychology suggest that the muscle analogy may in fact be an apt one. A number of researchers including Roy Baumeister, Todd Heatherton, Diane Tice, Kathleen Vohs and their colleagues have systematically examined a family of processes that they call 'self-regulatory processes', which are closely related to, but in many ways broader than the notion of willpower discussed in this article (Baumeister and Heatherton). (The relationship between self-regulatory processes and willpower will be taken up in the following section.) According to these researchers, self-regulatory processes can be characterized in terms of a 'strength' model (Baumeister et al.). The basic idea behind the model is that like the strength associated with a muscle, self-regulatory processes fatigue when exerted for prolonged periods of time. With rest, however, these processes regain their effectiveness. Over longer periods of time, the regular use of self-regulatory processes can lead to enhancements in their efficacy.

These researchers explored various aspects of the fatigability of self-regulatory processes by performing a number of experiments that have a characteristic structure (Fig. 1). Each experiment has two phases and the second phase occurs shortly after the first. In the first phase, one group of subjects is given a task that demands the use of self-regulatory processes. Examples of such tasks include inhibiting the tendency to read subtitles during a

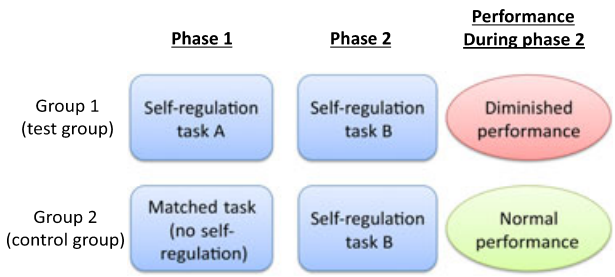

Fig. 1. Schematic structure of 'self-regulation experiments'. Results suggest that subjects in Group 1 *depleted* their self-regulatory capacities during the first task, thus leaving them less able to exert regulatory control during the second task.

movie, regulating one's emotions during a disturbing film, restraining the urge to eat a tempting food, suppressing thoughts with a certain content, resisting the urge to remove one's hand from a tank of extremely cold water, and maintaining one's concentration on a difficult- or impossible-to-solve puzzle. The second group of subjects is given a task matched in most respects to the task performed by the first group, but which does *not* demand the use of self-regulatory processes (e.g., subjects might watch the same film as the first group, but be allowed to read the subtitles). Both groups then perform a second task (distinct from the first task) that demands the use of self-regulatory processes. A robust and consistent finding across dozens of these experiments is that in the second phase of the experiment, the first group of subjects performs significantly worse than the second group of subjects. The authors' interpretation of this result is that the subjects in the first group *deplete* their self-regulatory capacities during the task in the first phase of the experiment, thus leaving them less able to exert regulatory control during the task in the second phase of the experiment.

If the strength model of willpower endorsed by these researchers is correct, then this would seem to put pressure on Feinberg's claim that the capacity to resist a temptation is limited only by one's motivation to resist, and never by fixed ceilings on the efficacy of the resistance mechanism. But we must be careful not to interpret the results of these 'self-regulation' experiments in ways that go beyond what the data can actually support. These self-regulation experiments only demonstrate that willpower has certain properties of a muscle – most importantly the property of exhibiting diminished efficacy immediately following sustained use. But if further research can deepen the analogy between willpower and muscular activity,[4] and in particular can provide direct evidence that willpower too exhibits fixed ceilings in its efficacy that hold independent of the motivation to exercise willpower, then Feinberg's claim would indeed be seriously challenged.

Question 3: How Many 'Willpowers' are there? What is the Relationship Between Willpower and Other Self-Regulatory Processes?

In the experiments by Baumeister and colleagues described in the previous section, researchers used the term 'self-regulatory process' to describe a family of processes that exhibit a common structure: these processes, once engaged, cause some 'target mental state' to be attenuated, blocked or rendered inefficacious. Willpower, narrowly construed, is a species of self-regulatory process where the target mental state is one of the agent's desires. Other self-regulatory processes attenuate or suppress a variety of other target mental states such as thoughts, emotions, urges, cravings, attentional distractions, and habitual or 'prepotent' responses.

A question then arises as to how willpower, understood narrowly as a capacity to inhibit one of the agent's own desires, relates to the broader family of self-regulatory processes. One hypothesis is that these various different kinds of self-regulatory processes are each implemented by largely distinct, although perhaps partially overlapping, neural mechanisms. One method for testing this hypothesis involves studying a variety of tasks, each one relatively selective in engaging just one kind of self-regulatory process. By studying multiple such tasks, it might be possible to 'parse' self-regulation into component processes.

For example, the *Stroop task* (MacLeod) is often used to probe *attention regulation* mechanisms involved in suppressing distractions. In this task (Fig. 2a), subjects are given a series of words and asked to state the ink color of the word. But the words themselves are names of colors, and the ink color and named color are often different. Hence, to per-

Fig. 2. Examples of tasks that are hypothesized to selectively engage distinct self-regulatory processes. (a) The Stroop Task in which subjects must state the ink color of the word rather than read the name of the color; (b) the Delay Discounting Task in which subjects choose between smaller, earlier versus larger, later monetary rewards; (c) a picture from an emotion regulation task in which subjects use 'distancing' strategies to attenuate responses to the emotional stimulus.

form successfully, subjects must avoid being distracted by the word's meaning and instead focus their attention on the ink color of the word. In the *delay discounting task* (Myerson and Green; Frederick, Loewenstein, and O'Donoghue), subjects are given choices between smaller, earlier or larger, later monetary rewards (Fig. 2b). One influential hypothesis holds that choices for later rewards depend on the engagement of *appetitive regulation* processes that suppress the 'default' tendency to choose more immediate rewards. In *emotion regulation tasks* (Ochsner et al.), subjects are confronted with emotionally salient stimuli, such as a highly disturbing picture, and asked to subjectively distance themselves from the picture (e.g., they might be asked to take a detached third-person perspective toward a frightening picture, Fig. 2c). This task is thought to engage processes specialized for attenuating responses to emotionally valenced stimuli. By systematically investigating performance in these three tasks, using a variety of methods such as reaction time, electrophysiology, or neuroimaging, it might be possible to determine interrelationships between the mechanisms implementing attention regulation, appetitive regulation, and emotion regulation. Overall, investigations aimed at parsing self-regulatory processing into component mechanisms (Friedman and Miyake; Ochsner and Gross; Wager et al.; Nee, Wager, and Jonides) are still at the early stages, but this area of research is one of the most active in cognitive neuroscience and promises to yield exciting results in the future.

Even if it is discovered that distinct neural mechanisms implement different forms of self-regulatory processing, these various and sundry mechanisms may nonetheless rely on a common energetic store. This hypothesis is supported by the self-regulatory experiments of Baumeister and colleagues. Recall that the range of tasks utilized in these experiments appears, at least superficially, to be quite heterogeneous. Nonetheless, subjects' engagement in one kind of task reliably leads to poorer performance on a subsequent task. If the processes engaged in these varied tasks all draw upon a common energetic store, one could naturally explain why prolonged engagement in any one of these tasks leads to subsequent poorer performance in any other.

Suppose that all self-regulatory processes exhibit a fixed ceiling on their efficacy, perhaps due to their reliance on a common energetic store (see Note 3), or perhaps for some other reason. This 'single ceiling thesis' calls into question certain arguments and distinctions that are frequently found in the philosophy and legal theory literature. For example, the legal theorist Stephen Morse has argued in a number of papers (Morse, 'Culpability and Control'; Morse, 'Hooked on Hype: Addiction and Responsibility; Morse, 'Uncontrollable Urges and Irrational People') that persons addicted to drugs *do not* deserve an excuse for actions in violation of the law based on *irresistible desires*. Roughly, his argu-

ment is that the idea that one's own desires 'internally coerce' one's behavior, once carefully examined, cannot be made sense of. However, more recently, Morse has argued that addicts *do* deserve an excuse, not due to irresistible desires, but rather due to the manner in which addiction impairs one's 'rational capacities'.

> …the addict, metaphorically, and in some cases perhaps literally, can think of nothing else but the desire to use the substance. One informant described the desire like "a buzzing in my ears that prevents me from focusing." … There is only one tune or story in the addict's head and nothing can drive it out… Fundamental components of rationality – the capacities to think clearly and self-consciously to evaluate one's conduct – are compromised (Morse, 'Hooked on Hype: Addiction and Responsibility' 39).

In the preceding paragraph, Morse seems to be suggesting that addiction excuses certain behaviors in virtue of its generating *irresistible thoughts*, i.e., the addict can think about nothing but obtaining the drug and these thoughts are incredibly hard to redirect or suppress. But it is not at all clear why Morse thinks irresistible thoughts and irresistible desires are on such different footings, so that the former are a proper basis for legal excuse but the latter are not. For one thing, given Morse's careful and exhaustive enumeration of reasons to doubt that one's own desires are ever truly irresistible, one wonders why these very same arguments don't apply to thoughts, thus showing that one's own thoughts are never truly outside of one's own control. A second potential problem for Morse is that his argument relies on an implicit distinction between the psychological mechanisms that enable one to control one's thoughts and attention versus those that enable one to control one's desires, such that impairments in the former are deserving of excuse but impairments in the latter are not. But it is unclear from Morse's writings in what this crucial difference consists. Indeed, if the 'single ceiling thesis' is correct that all self-regulatory processes (including processes that regulate desire as well as processes that regulate thought and attention) exhibit a common, fixed limit on their efficacy, then it stands to reason that these two categories of self-regulatory failure, although they may superficially *appear* quite different, should in fact be treated very much the same.

Morse's endorsement of mitigation for addicts based on impairments in rational capacities, but not on the basis of irresistible desires, reflects a widespread tendency for the law to treat irresistible impulses and other 'motivation-related' bases of excuse as less justified than putatively strictly 'cognitive' bases of excuse (Goldstein; Caplan). A fully fleshed out account of willpower that clarifies the relationship between willpower (which, per definition, is directed at regulating desires) and related processes directed at regulating thoughts and attention might call into question the strong preference for cognitive bases of excuse currently embraced in the law and in legal thinking.

Question 4: What is the Relationship Between Willpower and Weakness of Will (and Compulsion)?

On one well-accepted formulation, an agent's action is weak-willed if the agent freely and intentionally acts contrary to her all things considered judgment of what it would be best to do (Stroud and Tappolet). In a highly influential paper (Watson), Gary Watson raises skeptical questions about whether a weak-willed agent is in fact genuinely *able* to resist her contrary desires, and in doing so, Watson forges a close link between philosophical accounts of weakness of will and accounts of willpower. Watson presents three versions of a case in which a woman who ought not to drink because of some obligation, nonetheless drinks.

> (1) the reckless or self-indulgent case; (2) the weak case; and (3) the compulsive case. In (1), the woman knows what she is doing but accepts the consequences. Her choice is to get drunk or risk getting drunk. She acts in accordance with her judgment. In (2) the woman knowingly takes the drink contrary to her (conscious) better judgment; the explanation for this lack of self-control is that she is weak-willed. In (3), she knowingly takes the drink contrary to her better judgment, but she is a victim of a compulsive (irresistible) desire to drink (Watson 324).

Watson argues that according to the 'common sense' account of weakness of will, the weak case is like the compulsive case in that the agent acts contrary to her best judgment. But the weak case is like the reckless case in that the agent has the ability to resist, but fails to exercise it. But given that the weak agent is strongly (if not decisively) motivated to resist, this motivation arising from her all things considered best judgment, and given that she is able to resist, why does she fail in exercising resistance?

Watson considers a number of candidate answers to this question, such as that the weak agent does not want to go through the trouble of resisting, or that the weak agent underestimates the effort that would be needed to successfully resist. In each case, careful analysis finds the explanation clearly wanting. Since no basis can be found for why an agent who is both motivated and able to resist nonetheless fails to make the requisite effort at resistance, Watson argues that we are entitled to conclude that the common account of weakness of will is in fact critically mistaken – the weak-willed agent, at the time that she capitulated to her wayward desire, was in fact *unable* to resist (p. 338).

Watson's skeptical view is certainly controversial, and other writers (Buss; Mele, *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*; Tenenbaum; Mele, 'Akratics and Addicts') have reached an opposed conclusion. It will not be possible to discuss all the arguments and counterarguments in this debate. Instead, here I want to focus on the observation that there is a strong tendency to find the Watsonian skeptical position simply *unsatisfying* as a resolution to the question 'Is weakness of the will possible?', and the long list of authors opposed to Watson's conclusion attests to this claim. Moreover, it is likely that this dissatisfaction is rooted in the fact that common sense not only holds strongly to the idea that weak-willed actions are possible, but it also in fact insists that they routinely occur. But why might common sense be so insistent on the truth of these claims? Alfred Mele provides a succinct answer.

> Why do ordinary folks believe that there are (in this author's terminology, not theirs) strict akratic actions? Presumably, largely because they take themselves to have first hand experience of such action and partly because some of their observations of others indicate to them they are not alone in this… It is possible that these ordinary agents are wrong about this, of course… But why should one believe that they are wrong? (Mele, 'Akratics and Addicts' 159)

Mele suggests that one important source of the folk belief that weak-willed actions occur is 'first-hand experience'.[5] It is not exactly clear what Mele has in mind, but one plausible interpretation of 'first-hand experience' is in terms of the *phenomenology* associated with putatively weak-willed action. When a person caves in to temptation, the person experiences a distinctive suite of subjective experiences. For example, consider Charlie who is on a diet but feels tempted to eat a slice of cherry chocolate cake. As Charlie caves in to temptation, he experiences a characteristic phenomenology. His experiences typically include a *feeling of attraction* directed at the chocolate cake, a *feeling of effort* as he tries to resist eating the cake, and, eventually, the *feeling of giving in* to temptation. Most important for our purposes, as Charlie gives in to temptation, the resulting action is often accompanied by a '*feeling of uncompelledness*'. This feeling is difficult to articulate, but very roughly, Charlie feels the action is not forced on him by a source external to him.

Rather, he feels that he is the author of the action, and the action is a consequence of his choosing. So one way of understanding Mele's claim that the folk have 'first-hand experience' of weak-willed action is that people often have a feeling of uncompelledness when they undertake actions that contravene their best judgment. This feeling 'depicts' their action to themselves as free and intentional. In this way, people come to believe that there are weak-willed actions.

I think there is something right about the proposal that the 'feeling of uncompelledness' is the central underpinning for the folk belief that weak-willed actions routinely occur. But if this is so, then it points to an important goal for philosophical research into the nature of willpower. In particular, there is a pressing need to answer the question of whether the feeling of uncompelledness that routinely accompanies certain kinds of failure of willpower is *veridical*. That is, when an agent experiences a feeling of uncompelledness as she performs an action that contravenes her best judgment, is this feeling of uncompelledness accurate in the way that it represents the agent as acting freely, intentionally, and without compulsion?

One place to look in beginning to answer this question is the growing literature in philosophy and neuroscience about the subjective experience of willed action (Bayne and Levy; Bayne; Haggard). This literature is beginning to illuminate some of the brain mechanisms by which various experiences associated with agency (such as the experience of *authorship*, and the feeling of *doing*) become 'attached' to actions. A key lesson from this literature is that there are multiple ways in which the mechanisms that link actions with authorship experiences can misfire, so that some actions that the agent in fact authors are *not* tagged with authorial experiences, while other actions the agent does not fully author *are* inappropriately tagged. The feeling of uncompelledness associated with weak-willed actions may simply *be* the feeling of authorship. Or it may be a complex of experiences in which authorship experiences are only a component. But a better understanding of the brain mechanisms by which the feeling of uncompelledness is generated will be crucial to understanding the component structure, if any, of this experience, as well as whether, and under what conditions, the experience may fail to be fully veridical in depicting the agent as originating and authoring an action.[6]

Suppose further investigation revealed that in paradigm cases in which an agent acts in a putatively weak-willed manner, the feeling of uncompelledness is not veridical because the agent is in some important sense compelled to act as she does. This sort of finding would provide critical support for skepticism about weakness of will. A central reason for dissatisfaction with the skeptical position that weak-willed actions do not really exist is, as Mele correctly points out, that we seem to be acquainted with the fact of their existence through first-person experience. Thus, an account of why first-person experience might be systematically mistaken about the existence of weak-willed actions would do much to loosen the grip of common sense, and thereby would significantly temper the dissatisfaction with which the skeptical position is typically greeted.

Conclusion

In this article, I surveyed four key questions about willpower: How is willpower possible? Why does willpower fail? How does willpower relate to other self-regulatory processes? and What are the connections between willpower and weakness of will? As we have seen, the topic of willpower is closely bound up with a number of important philosophical questions about human agency. While there is much we still do not know about willpower, there is also justified excitement these days that as empirical investigations

progress, we are coming ever closer to reaching satisfying answers to at least some of the key questions posed in this article.

Acknowledgements

Earlier versions of parts of this article were presented at the University of Michigan Decision Consortium Group, the Moral Psychology Research Group (Pittsburgh, PA), and the 'Philosophy and Science of Weakness of Will' series at Amherst College. I am grateful to these audiences for their comments. Special thanks to Sarah Buss who provided invaluable comments on practically every aspect of this manuscript.

Short Biography

Chandra Sekhar Sripada works on issues about human agency and action that lie at the intersection of philosophy and the behavioral and brain sciences. He received his PhD in philosophy from Rutgers University and completed residency training in psychiatry at the University of Michigan. He currently holds a joint appointment at the University of Michigan. One appointment is in the Department of Philosophy where he works mainly in metaethics and moral psychology. The other appointment is in the Department of Psychiatry, where he studies the neuroscience of decision-making and self-control, and breakdowns in these capacities associated with psychiatric disorders. He is currently authoring a set of papers on the concepts of free will, moral responsibility, intentional action, and mental disorder, with the aim of illuminating underlying unities among these concepts.

Notes

⋆ Correspondence: Departments of Philosophy, University of Michigan, 2215 Angell Hall, 435 South State Street, Ann Arbor, MI 48109, USA. Email:sripada@umich.edu.

[1] I thank Nishiten Shah for the helpful idea that the puzzle of synchronic self-control should be formulated as a dilemma.

[2] Dual-process models are one of a number of empirically supported frameworks available in cognitive psychology that postulate that information processing is implemented by multiple-interacting psychological compartments. In what follows, I illustrate the divided mind view of willpower using a dual-process framework, but recognize that alternative frameworks for partitioning the mind might have also been used.

[3] I develop a divided mind account of synchronic self-control and address these, as well as other questions raised by the account, in 'The puzzle of resistance and the divided mind'.

[4] Recent studies suggest that self-regulatory processes are fueled by glucose (Gailliot and Baumeister), and prolonged engagement of these processes depletes glucose (Gailliot et al.; Masicampo and Baumeister). These studies are tentative and await corroboration, but they are a good example of the type of findings that would help deepen the analogy between willpower and muscular exertion, and would help support the view that there are fixed ceilings in the efficacy of willpower that arise independent of one's motivation to resist.

[5] I use the term 'weak-willed' in place of Mele's term 'akratic' in what follows.

[6] In 'An error theory of weakness of will', I more fully develop an account that holds that weak-willed actions do not exist, and that proposes that the error in common sense in believing they do exist arises from, *inter alia*, the misleading subjective experiences characteristically associated with exercises of willpower.

Works Cited

Baumeister, R. F. and T. F. Heatherton. 'Self-Regulation Failure: An Overview.' *Psychological Inquiry* 7.1 (1996): 1–15.

——, —— and D. M. Tice. *Losing Control: How and Why People Fail at Self-Regulation*. San Diego: Academic Press, 1994.

Baumeister, Roy. E., Ellen Bratslavsky, Mark Muraven and Diane M. Tice. 'Ego Depletion: Is the Active Self a Limited Resource?.' *Journal of Personality and Social Psychology* 74.5 (1998): 1252–65.
Bayne, Timothy. 'The Phenomenology of Agency.' *Philosophy Compass* 3.1 (2007): 182–202.
—— and N. Levy. 'The Feeling of Doing: Deconstructing the Phenomenology of Agency.' *Disorders of Volition*. Ed. N. Prinz and W. Sebanz. Cambridge, MA: MIT Press, 2006. 49–68.
Bechara, A. 'Decision Making, Impulse Control and Loss of Willpower to Resist Drugs: A Neurocognitive Perspective.' *Nature Neuroscience* 8.11 (2005): 1458–63.
Beck, Aaron T. *Cognitive Therapy and the Emotional Disorders*. New York: International University Press, 1976.
——, A. J. Rush, B. F. Shaw and G. Emery. *Cognitive Therapy of Depression*. New York: Guilford Press, 1979.
Buss, Sarah. 'Weakness of Will.' *Pacific Philosophical Quarterly* 78.1 (1997): 13–44.
Caplan, L. *Insanity Defense and the Trial of John W Hinckley, Jr.* Boston: David R Godine Publishers, Inc, 1984.
Chaiken, S., A. Liberman and A. H. Eagly. 'Heuristics and Systematic Processing Within and Beyond the Persuasion Context.' *Unintended Thought*. Eds. J. S. Uleman and J. A. Bargh. New York: Guilford Press, 1989. 212–52.
Davidson, Donald. 'Paradoxes of Irrationality.' *Philosophical Essays on Freud*. Eds. Richard. Wollheim and James Hopkins. Cambridge: Cambridge University Press, 1982. 289–305.
Feinberg, Joel. 'What is so Special About Mental Illness?' *Doing and Deserving*. Princeton: Princeton University Press, 1970. 272–92.
Frederick, S., G. Loewenstein and T. O'Donoghue. 'Time Discounting and Time Preference: A Critical Review.' *Journal of Economic Literature* 15 (2002): 351–401.
Friedman, N. P. and A. Miyake. 'The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis.' *Journal of Experimental Psychology: General* 133.1 (2004): 101–35.
Gailliot, Matthew T. and Roy F. Baumeister. 'The Physiology of Willpower: Linking Blood Glucose to Self-Control.' *Personality and Social Psychology Review* 11.4 (2007): 303–27.
Gailliot, Matthew T., C. Nathan DeWall, Jon K. Maner, E. Ashby Plant, Dianne M. Tice, Lauren E. Brewer and Brandon J. Schmeichel. 'Self-Control Relies on Glucose as a Limited Energy Source: Willpower is More Than a Metaphor.' *Journal of Personality and Social Psychology* 92.2 (2007): 325–36.
Gilbert, Daniel. 'What the Mind's Not.' *Dual-Process Theories in Social Psychology*. Eds. S. Chaiken and Y. Trope. New York: Guilford Press, 1999. 3–11.
Gilovich, T., D. Griffin and D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, MA: Cambridge University Press, 2002.
Goldstein, Abraham. *The Insanity Defense*. New Haven: Yale University Press, 1967.
Haggard, P. 'Human Volition: Towards a Neuroscience of Will.' *Nature reviews Neuroscience* 9.12 (2008): 934–46.
Hofmann, Wilhelm, Malte Friese and Fritz Strack. 'Impulse and Self-Control From a Dual-Systems Perspective.' *Perspectives on Psychological Science* 4.2 (2009): 162–76.
Holton, Richard. *Willing, Waiting, Wanting*. New York: Oxford University Press, 2009.
Kahneman, D. 'A Perspective on Judgment and Choice: Mapping Bounded Rationality.' *American Psychologist* 58 (2003): 697–720.
Kennett, Jeanette and Michael Smith. 'Frog and Toad Lose Control.' *Analysis* 56.2 (1996): 63–73.
—— and ——. 'Synchronic Self-Control Is Always Non-Actional.' *Analysis* 57.2 (1997): 123–31.
Loewenstein, George. 'Willpower: A Decision-Theorist's Perspective.' *Law and Philosophy* 19.1 (2000): 51–76.
MacLeod, Colin M. 'Half a Century of Research on the Stroop Effect: An Integrative Review.' *Psychological Bulletin* 109.2 (1991): 163–203.
Masicampo, E. J. and Roy F. Baumeister. 'Toward a Physiology of Dual-Process Reasoning and Judgment: Lemonade, Willpower, and Expensive Rule-Based Analysis.' *Psychological Science* 19.3 (2008): 255–60.
Mele, Alfred. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press, 1992.
——. 'Akratics and Addicts.' *American Philosophical Quarterly* 39.2 (2002): 153–67.
——. *Motivation and Agency*. New York: Oxford University Press, 2003.
Metcalfe, J. and W. Mischel. 'A hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower.' *Psychological Review* 106.1 (1999): 3–19.
Morse, Stephen J. 'Culpability and Control.' *Pennsylvania Law Review* 142 (1994): 1587–660.
—— 'Hooked on Hype: Addiction and Responsibility.' *Law and Philosophy* 19.1 (2000): 3–49.
—— 'Uncontrollable Urges and Irrational People.' *Virginia Law Review* 88.5 (2002): 1025–78.
Myerson, J. and L. Green. 'Discounting of Delayed Rewards: Models of Individual Choice.' *Journal of the Experimental Analysis of Behavior* 64.3 (1995): 263–76.
Nee, D. E., T. D. Wager and J. Jonides. 'Interference Resolution: Insights From a Meta-Analysis of Neuroimaging Tasks.' *Cogn Affect Behav Neurosci* 7.1 (2007): 1–17.
Ochsner, Kevin N. and James J. Gross. 'The Cognitive Control of Emotion.' *Trends in Cognitive Sciences* 9.5 (2005): 242–9.

Ochsner, K. N., R. D. Ray, J. C. Cooper, E. R. Robertson, S. Chopra, J. D. Gabrieli and J. J. Gross. 'For Better or for Worse: Neural Systems Supporting the Cognitive Down- and up-Regulation of Negative Emotion.' *Neuroimage* 23.2 (2004): 483–99.

Sanfey, Alan G., George Loewenstein, Samuel M. McClure and Jonathan D. Cohen. 'Neuroeconomics: Cross-Currents in Research on Decision-Making.' *Trends in Cognitive Sciences* 10.3 (2006): 108–16.

Sloman, S. A. 'The Empirical Case for two Systems of Reasoning.' *Psychological Bulletin* 119 (1996): 2–22.

Stanovich, K. E. and R. F. West. 'Individual Differences in Reasoning: Implications for the Rationality Debate.' *Behavioral and Brain Sciences* 23 (2000): 645–65.

Stroud, S. and C. Tappolet. *Weakness of Will and Practical Irrationality*. New York: Oxford University Press, 2008.

Tenenbaum, Sergio. 'The Judgment of a Weak Will.' *Philosophy and Phenomenological Research* 49.4 (1999): 875–911.

Wager, T. D., C. Y. Sylvester, S. C. Lacey, D. E. Nee, M. Franklin and J. Jonides. 'Common and Unique Components of Response Inhibition Revealed by FMRI.' *Neuroimage* 27.2 (2005): 323–40.

Watson, Gary. 'Skepticism About Weakness of the Will.' *Philosophical Review* 86 (1977): 316–39.