

# Freedom and the Will

Kieran Setiya

MIT

February 13, 2024

(Draft; do not cite without permission)

*Freedom within Reason* (Wolf 1990) is a classic in the contemporary debate about freedom and responsibility. It is short but characteristically bold. In it, Susan Wolf argues against the most influential forms of compatibilism and incompatibilism, develops her own distinctive Reason View, defends the asymmetries it generates—between being unable to do the wrong thing, which is consistent with freedom and responsibility, and being unable to do the right thing, which is not—and uses a thought-experiment about God’s creation to argue that physical as opposed to psychological determinism is no threat to culpability for wrongdoing. Wolf was inspired by two brief but radical essays: P. F. Strawson’s “Freedom and Resentment” (Strawson 1962) and Iris Murdoch’s book, *The Sovereignty of Good* (Murdoch 1970).<sup>1</sup>

My project is less ambitious. I won’t say anything about responsibility or blame—concepts I find baffling, blurring as they do the metaphysics of freedom, the nature of such “reactive attitudes” as resentment and indignation, and social practices of holding one another accountable. Culpability is, as Gideon Rosen once remarked, “the dark underbelly of moral philosophy”; and I am afraid of the dark. Instead, I will explore another brief but radical intervention, one that has had less impact, perhaps because of its ironic and elusive style: Rogers Albritton’s APA address, “Freedom of Will and Freedom of Action” (Albritton 1985). My aim is to criticize an assumption taken for granted by almost everyone who writes about freedom in the philosophy of action, including Wolf: that cases of compulsion involve a deficit of freedom that distinctively afflicts the will. In making my case, I extrapolate from Albritton’s telegraphic notes, sketching an account of the ability to act, along with a

---

<sup>1</sup> See her Preface (Wolf 1990: ix).

partial theory of strength and weakness of will. I end with a deflationary argument for Descartes' thesis that "the will is perfectly free, 'so free in its nature that it cannot be constrained'" (Albritton 1985: 239).

## I

Albritton, too, defends Descartes' position, perhaps with tongue in cheek. What he argues, more explicitly, is that there is no circumstance in which we can truly say that someone "lacks freedom of the will": we cannot make good sense of that description. It isn't obvious why he doesn't go "full Wittgenstein," holding that, since the denial of free will makes no sense, the assertion of it, too, is nonsense.<sup>2</sup> Instead, he is gnominically assertoric: "free will is what we've got if the will is free, as of course it is," he insists. "I don't exactly know what any of that means, but I don't know how to doubt that we've got free will, either" (Albritton 1985: 243).

It is tempting just to quote the wry remarks that leaven the essay—"No doubt we're free as birds. ... But how free *are* birds?"—or the moments of snark, as when we meet "the horrid tribe of 'soft determinists'" (Albritton 1985: 239, 240). But although it rambles, the structure of Albritton's argument is clear. It begins by distinguishing a loss of freedom to act from a deficiency of free will. It then considers, in turn, alleged impediments to free will—epistemic, normative, psychological, deterministic—arguing that each is misconceived or misdescribed. I will concentrate on psychological incapacity, where Albritton's best insight lies. But it is useful to locate this insight in the broader argument of the essay, marking points of disagreement.

Like many, I agree that we should distinguish freedom of action from freedom of will, at least in admitting impediments to the former that do not affect the latter. Albritton begins by citing Anscombe to the contrary: "everyone will allow," she claims, that "a physical impossibility of, say, walking does restrict one's freedom of will *if* it comes of some external constraint" (Albritton 1985: 240). "I don't allow it," is Albritton's blunt response (Albritton 1985: 240). Being put in chains restricts

---

<sup>2</sup> Compare his posthumously published essay on scepticism and epistemic possibility (Albritton 2011).

one's freedom or ability to walk—but, as Albritton complains, there is no good sense in which it removes one's "freedom of the will in respect of walking" (Albritton 1985: 240). Chains "diminish my already unimpressive capacity to *do* what I will," he writes, but not the freedom of my will (Albritton 1985: 240). To underline the point, Albritton imagines trying to go for a walk, only to find himself in chains: "Do I have reason to think not only, 'They've chained me up!' but, 'Good God, they've been tampering with my will!?' No, I don't" (Albritton 1985: 241). The same is true of the "curare of [Albritton's] imagination," which induces total physical paralysis. By itself, it is no hindrance to the *will*: though it prevents me from moving, it does not prevent me from intending or trying to move (Albritton 1985: 243).

This all seems right to me, and relatively uncontentious: with Albritton, we should distinguish "something obscure that might reasonably be called 'freedom of will' and every kind of freedom to *work* one's will, to *do* as one will, to *have* one's will or way, so to speak" (Albritton 1985: 242). More contentious is what Albritton goes on to say about the epistemology of the will. What if I know that I am unable to  $\phi$ ? Does that limit my freedom of will? Again, he thinks not:

It isn't that my will is hobbled by the prospect of an impossible project. It's that 'I know I can't do it, but I've decided to do it anyway,' for example, is either a figurative way of speaking or a kind of nonsense. Perhaps it is even a contradiction. ... what gets in the way of *deciding* to walk when you know you can't, or think you know you can't, is in the language, not in the will. (Albritton 1985: 244)

Here I disagree—I think. If it's impossible to intend to  $\phi$  when I believe that I am unable to  $\phi$ , the obstacle to doing so is metaphysical, not linguistic. (You can't explain this necessity by appeal to facts about language.) And if intention involves belief, or some degree of belief, that one will  $\phi$ , it may be epistemically irrational—though possible—to decide, and so intend, to do what you know you can't.<sup>3</sup>

---

<sup>3</sup> Here, and throughout, I equate deciding with forming an intention. On the relationship between intention and belief, see Setiya 2016.

But again, the obstacle is not linguistic. I am thus willing to concede that there may be metaphysical or epistemic limits on the will. If Albritton denies this, he is wrong. The crucial point, however, is that these are not the sorts of limits at issue in debates about freedom of the will, where the constraints are causal or psychological. I propose to set the metaphysics and epistemology of intention to one side.

Following Albritton, I set aside, as well, the normative sense of “must” in which I must do what I am obligated to. “I have to make my rounds, that’s true,” he writes, “and it follows, trivially, that I can’t stay. But I don’t have to make my rounds in any sense that binds my will. Even Luther’s reasons were only *reasons*” (Albritton 1985: 247). From the fact that I am obligated to  $\phi$ , it doesn’t follow that I am unable to do otherwise, let alone that I lack the freedom of will to do so.<sup>4</sup> If Luther lacked freedom of will when he said “Here I can stand; I can do no other” that was not because he was subject to normative necessity but because of the psychological incapacity it induced. It’s psychological incapacity that threatens our freedom of will.

## II

Albritton makes us wait until the final pages of his essay before turning to the sort of case that his opponents will have had in mind from the beginning. “I must say at least something about desire and aversion, and all that lot. That is, I am determined to,” he quips. “But it won’t be nearly enough” (Albritton 1985: 248). My aim is to say some of what Albritton does not.

What he does say is brief and breezy. About someone addicted to alcohol: “It’s simply not true, locally, that the alcoholic can’t help himself. I mean: he could perfectly well, and sometimes does, empty the bottle into the sink. But usually he doesn’t. He drinks it in (as far as I can see) full freedom of will” (Albritton 1985: 249). But suppose he *can’t* resist, on this occasion? “Perhaps he just hasn’t the strength of will to hold out, as one might be unable to withstand torture,” concedes Albritton. “But

---

<sup>4</sup> I may lack the freedom to do so in that I have no right to do otherwise, i.e., no right against others not to be simply made to do what I am obligated to do (Setiya 2022). But that is a different matter.

*strength* of will is one thing and *freedom* of will is another. Isn't it? Or do you think not? I think it is" (Albritton 1985: 249).

Not much of an argument, you might think. Or do you think otherwise? Certainly, it runs against a widespread view, according to which compulsion impedes freedom of action by subverting one's freedom of will. Here is Keith Lehrer in a much-cited paper:

Suppose that I am offered a bowl of candy and in the bowl are round red sugar balls. I do not choose to take one of the red sugar balls because I have a pathological aversion to such candy. ... It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but, not so choosing, I am utterly unable to touch one. ... I could do it only if I chose to, and I do not. (Lehrer 1968: 32).

This is taken to refute a simple conditional analysis of freedom: S is able to  $\phi$  just in case if S intended to  $\phi$ , S would  $\phi$ . Lehrer's candyphobe is, allegedly, unable to take a red sugar ball, even though he would have taken one had he intended to, because he can't form that intention. What he lacks is freedom of will.

Wolf agrees, comparing kleptomania to hypnosis and coercion: "in each case the problem is neither with the effectiveness nor with the content of the agent's will. It is rather with the source of the agent's will—with the fact that the agent is not in control of what the content of his or her will will be" (Wolf 1990: 10). She states the consensus as follows:

We have already agreed that this agent is at liberty to perform whatever action she wills to perform—what she thinks she lacks, and what she may truly lack, is the liberty to will whatever actions she wants to will. She is alienated from her will, because her will is not, in an important sense, the result of her choice. (Wolf 1990: 29)

It's a consensus that goes back to David Hume, who defends a conditional analysis of freedom, but agrees with Lehrer and Wolf that its conditions are met in the absence of external constraints:

By liberty, then, we can only mean *a power of acting or not acting, according to the determinations of the will*; that is, if we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty is universally allowed to belong to every one, who is not a prisoner and in chains. (Hume 1748: 8.23)

But to echo Albritton: I don't allow it.

What happens to the compulsive described by Albritton is that he intends not to drink (or not to confess under torture) and fails to abide by this intention. He lacks the ability *not* to drink (or to withstand torture) because he lacks the ability to execute his intention, not the ability to form it. There is no need to posit any deficit of free will, here, whatever that would mean. Likewise, with Lehrer's candyphobe, realistically conceived. If you tell me that a man is pathologically averse to red sugar balls, I'll infer that, even if he set himself to take one, he wouldn't succeed. His pathology would intervene and he would fail to act on his intention. He is thus deprived of ability in the conditional sense. If you reply by stipulating that it isn't so—his aversion ensures that he won't decide to take a red sugar ball, but if he did decide that, he would take one—I'll conclude that he is perfectly able to take one, just determined not to. The same is true if you add that the candyphobe *can't* intend to take a red sugar ball, in that it's circumstantially impossible: his psychology prevents it. That doesn't mean he's unable to act on that intention. Nothing could induce me to run a marathon: I would never decide to do that, under any circumstance; it's not a possibility. But I am pretty sure I am able to run a marathon, if "running" very slowly counts. In each case, the simple conditional analysis gets things right.

When Albritton jokes that "*strength* of will is one thing and *freedom* of will is another," his point is that strength of will is exercised downstream of intention: it has to do with acting on one's intention in the teeth of motivational pressure (Albritton 1985: 249). Conversely, weakness of will, whatever else it may involve, is a matter of revising or failing to execute one's intention. Freedom of

will, by contrast, is meant to be about what happens upstream of intention, before one's intention is formed. But in ordinary cases, the compulsive agent's struggle is a struggle to resist temptation, a battle between strength and weakness of will. It's about the freedom to act as one wills, not freedom of will, though the obstacles to action are not chains but aspects of one's own psychology.

In "Volitional Necessities," Gary Watson quotes a passage from an essay by Andrew Solomon on the disabling effects of depression:

I ran home shaking and went to bed, but I did not sleep, and could not get up the following day. I wanted to call people to cancel birthday plans, but I couldn't. ... I knew that for years I had taken a shower every day. Hoping that someone else could open the bathroom door, I would, with all the force in my body, sit up; turn and put my feet on the floor, and then feel so incapacitated and frightened that I would roll over and lie face down. I would cry again, weeping because the fact that I could not do it seemed so idiotic to me. (Solomon 1998: 46, 49; quoted in Watson 2002: 93)

Here is Watson's commentary:

Solomon struggles hard to sit up in an effort to get out of bed; he sees and responds to reasons to arise. But his fear overcomes him. His fear is an obstacle to his efforts to carry out his intentions. It prevents him from carrying out his will. So this is just incapacity in the basic sense, it seems. (Watson 2002: 94)

Just so: when we turn to realistic cases of infirmity of the will, the infirmity lies in the inability to execute one's intentions. Yet Watson continues: "This analysis seems wrong. It doesn't deal with cases in which the agent is too terrified even to try" (Watson 2002: 94). He doesn't offer an example of this, perhaps because examples like Lehrer's are so familiar. But as I've emphasized, I don't find Lehrer's reading of the case, or cases like it, compelling.

We are left, I think, without a clear instance of psychological incapacity in which an agent lacks the freedom to act intentionally because she lacks freedom of the will. On the other hand, we are also left without a clear conception of ability—can the simple conditional analysis be right?—or of what is involved in strength and weakness of will. Only in the light of such conceptions can we verify Albritton’s hunch.

### III

To say that I am able to  $\phi$  is not just to say that I could. Although I am terrible at darts, I might hit a bullseye with this one; it’s not impossible. But if I say “I can hit a bullseye with this dart,” the salient reading is one on which I am guilty of false bravado.<sup>5</sup> Linguists and philosophers of language are thus led to distinguish “possibility modals,” which tell us what could or might occur from “ability modals,” which imply some sort of efficacy or control. The simple conditional analysis of ability seems to capture this element, and it gets the example right: it is not true that, if I intended to hit the bullseye with this dart, I would; so it is not true that I am able to do so.

But the analysis founders on inability.<sup>6</sup> In front of me is a ticking time bomb, presenting an array of coloured wires. I know that cutting one of the wires will defuse the bomb, though I’m not sure which one. I think it’s the green wire; but, unfortunately, I am wrong. Only the red wire will defuse the bomb; cutting any other wire will cause the bomb to detonate, which it will do in sixty seconds anyway. It is unclear, for reasons I will come to, whether I am able to defuse the bomb. But it’s not true that I am unable to, even though, if I decided to defuse the bomb, I would cut the green wire and thus fail. If someone said “He doesn’t know how to defuse that bomb,” they would be right. But if they said “He can’t defuse that bomb,” the most salient reading is one in which they would be saying something false:

---

<sup>5</sup> See Mandelkern, Schultheis, and Boylan 2017: 304; the discussion that follows is indebted to this paper, and to Boylan 2022.

<sup>6</sup> Compare Mandelkern, Schultheis, and Boylan 2017: 310 on what one “cannot not do” or “cannot but do.” I focus on the simpler case in which one cannot  $\phi$ .



that, whatever I decide to do, the bomb will detonate. I won't defuse the bomb—I am not going to—but that doesn't mean I can't. To say that someone is unable to  $\phi$  is to imply that, whatever they intend, they will not  $\phi$ . That's not my situation. If cutting the red wire will defuse the bomb, and were I to intend to cut that wire, I would, I am not unable to defuse it. More strongly, for intentional agents:

INABILITY: "S cannot  $\phi$ " is true just in case, for all  $\psi$ , if S intended to  $\psi$ , S would not  $\phi$ .<sup>7</sup>

(Note: this principle concerns "specific" ascriptions of (in)ability, the (in)ability to do something in the circumstance one occupies. There are also "general" ascriptions of ability, as when you say "I can swim" when you are miles from water. More on this below.)

Do we need the detour through " $\psi$ " in the case of ability, as well as inability? Yes, as J. L. Austin's notorious golfer shows. From a footnote in "Ifs and Cans":

Consider the case where I miss a very short putt and kick myself because I could have holed it. It is not that I should have holed it if I had tried: I did try, and missed. It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it. There is the rub.  
(Austin 1956: 218n)

Why is it true to say that I was able to make the putt, even though I intended to but didn't? Because I would have made it had I aimed a little differently, perhaps a touch to the left. There is a  $\psi$  such that, if I had intended to  $\psi$ , I would have succeeded in making the putt.<sup>8</sup> If there is no such  $\psi$ —if there is no

---

<sup>7</sup> A tricky case for Inability is that of the coma patient who cannot get out of bed. A confounding factor here is the temptation to suppose that, if S were to intend to get out of bed, S must have woken up from their coma: in the "closest world" in which they form this intention, they are no longer comatose. But the relevant question is whether having this intention, without any prior change in their condition, would cause them to get up. And the answer to that question is no: given their comatose state, their intentions fail to govern what they do.

<sup>8</sup> For a similar point, in terms of trying, see Chisholm 1964: 23-4. (As Chisholm observes, the detour through  $\psi$  solves Austin's other problems, too.) Chisholm goes on to reject his own proposal on the ground that I

intention about how to aim the putt on which it would have gone in—then it is not true to say that I was able to make the putt, even if it's true that I might have made it.

Turning this point into a biconditional is tricky because there seem to be contextual or pragmatic factors in the range of the variable “ $\psi$ .”<sup>9</sup> In the bomb defusal case, it seems wrong to say “I can defuse the bomb” since I don't know how. On the other hand, an observer who knew about the red wire would be in his rights to say “He can too defuse the bomb; he just needs to cut the red wire, and he can certainly do that.” Sometimes “ $\psi$ ” seems to be restricted to means I know will be effective in  $\phi$ -ing or by my conception of possible means; at other times, it seems unrestricted. Austin's golfer lies somewhere in between. In order for him to say truthfully “I was able to make the putt,” the relevant means must be ones he might have considered or to which he gave some credence. If the decision he would have had to make to sink the putt is one that would never have occurred to him, the truth of his assertion is in doubt. Without explaining the messy linguistic data, we can state two conditionals for ability, geared to intentional action:

ABILITY: “S can  $\phi$ ” is true only if there is some  $\psi$  such that, if S intended to  $\psi$ , S would  $\phi$ ; it is true if that condition is met, S knows that  $\psi$ -ing is a means by which to  $\phi$ , and S knows how to  $\psi$ .<sup>10</sup>

---

may be unable to  $\phi$ , even though I would  $\phi$  if I tried to  $\psi$ , because I am unable to try (Chisholm 1964: 24-5). He is right about this—it may be that, whatever I intend, I won't even *try* to  $\psi$ , so I am unable to  $\phi$ —but the problem goes away when we switch from the act of trying to intention as a mental state. To complain that, even though I'd  $\phi$  if I intended to  $\psi$ , I am not able to  $\phi$ , since I can't intend to  $\psi$ , is to return to the disputed case of Lehrer's candyphobe, discussed in the text below.

<sup>9</sup> For discussion, see Mandelkern, Schultheis, and Boylan 2017: 311-23. Another tricky phenomenon is the apparent implication from “S was able to  $\phi$ ” to “S  $\phi$ -ed” (Bhatt 1999). I am inclined to think this implication is pragmatic: it can be cancelled, as in the case of Austin's golfer, who was able to make a putt he missed.

<sup>10</sup> What about the abilities of inanimate objects, incapable of acting intentionally? I am not convinced that they exist. For data indicating that ascriptions of such “abilities” tacitly invoke intentional agents or are disguised possibility modals, see Mandelkern ms.: §6.4. In brief: for intentional ability claims, probabilities track the probability of success conditional on the agent's intending (or trying); for other alleged ability claims, probabilities either track the probability of success conditional on some relevant, perhaps generic, agent's intending (or trying) or they track the probability that a given event could or might occur, indicated that this is not a true ability modal. (Note that non-human animals count as intentional agents by the lights of Ability: the sense of “intention” here is not specific to rational animals like us.)

Where there is no  $\psi$  such that, if S intended to  $\psi$ , S would  $\phi$ , but it is not the case that, for all  $\psi$ , if S intended to  $\psi$ , S would not  $\phi$ , it is indeterminate whether S can  $\phi$ . That is how it is with me in the dartboard case: it is neither true nor false that I can hit a bullseye with this dart.<sup>11</sup> (Like Inability, the principle above concerns specific ascriptions, bound to particular circumstances. When you say “I can swim” while being miles from water, the reading on which your claim is true is general: it involves a generic claim about abilities in the sense at issue here.)<sup>12</sup>

The upshot of Ability and Inability is a partial, not-so-simple conditional analysis of freedom, one that vindicates my Albritton-adjacent take on the cases of compulsion and Lehrer’s candyphobe. The person who is unable to resist another drink, or to withstand torture, lacks the ability to do so because, whatever she intends, she won’t refrain from drinking, or confessing under torture. She needs help. Likewise, if Luther is unable to do otherwise, that is because, if he decided to recant, no matter how, the words would stick in his throat.

By contrast, if Lehrer’s candyphobe would take a red sugar pill if only he decided to—it’s just that he won’t, or can’t, make that decision—then he is able to take a red sugar pill. That’s not how I imagine the case when I am told that his aversion is pathological. Instead, I assume that, if he decided to take a red sugar pill, he would fail to act on that decision: his shaking hand would stall as he reached for the candy, stymied by disgust—or he would simply change his mind.

What is more, we can now add: there is no decision that would suffice to overcome his fear, no  $\psi$  such that, if he just decided to  $\psi$ , he would successfully take a red sugar ball. His problem is not that he doesn’t know how to get himself to do it, a problem he might solve with ingenuity—but that his phobia prevents even his best intentions. Whatever he decides right now, he won’t succeed in taking a red sugar ball.

---

<sup>11</sup> That we need to allow for indeterminacy in cases of this kind is one of the lessons of Boylan 2022. But his account predicts, I think mistakenly, that Lehrer’s candyphobe cannot even try to take a red sugar pill (Boylan 2022: §7). Note that, if “S can  $\phi$ ” is indeterminate when “S cannot  $\phi$ ” is determinately false, we get failures of “agentive duality” for abilities to act; see Loets and Zakkou 2022: §4.2.

<sup>12</sup> See Mandelkern, Schultheis, and Boylan 2017: §6.

This addition points to our second question: having said this much about ability and inability, can we explain what is involved in strength and weakness of will?

#### IV

In *Willing, Wanting, Waiting* (Holton 2009), Richard Holton gives an empirically informed account of willpower, distinguishing *akrasia* from weakness of will. In *akrasia*, one acts against one's better judgement. But one can act against one's better judgement with perfect resolution, displaying no weakness of will. And one can resolve to act against one's better judgement but be unable to maintain one's resolution, lacking the strength of will to do what one believes, or even knows, one shouldn't do. Defining a resolution as an intention to  $\phi$  conjoined with an intention not to let one's intention to  $\phi$  be deflected by temptation, Holton understands weakness of will as "unreasonable revision of a contrary inclination defeating intention (a resolution) in response to the pressure of those very inclinations" (Holton 2009: 11, 78).

As he goes on to document, temptation typically involves a shift in evaluative judgement, not a conflict between judgement and intention, on the one side, and desire or inclination, on the other.<sup>13</sup> In the grip of temptation, abandoning one's resolution seems like a good idea. Addiction is different. In the addict, motivation is decoupled from pleasure or enjoyment, and from evaluative judgement. What Albritton glibly remarked of the alcoholic, Holton's evidence confirms: real-life addicts are sometimes able to refrain from addictive behaviour, relying on willpower or strategies of self-control.<sup>14</sup> This generalization is consistent with the claim that, in particular cases, or on particular occasions, they are unable resist, like the compulsive imagined at the end of section III. In such cases, they will end up acting on their addiction, whatever they intend.

---

<sup>13</sup> This paragraph draws on Holton 2009: Ch. 5.

<sup>14</sup> See also Pickard 2015 on the clinical evidence of addicts' ability to do otherwise, though I would resist her suggestion that this ability is a condition of intentional action.

Holton's understanding of weakness is disputable. Unreasonable revision of a resolution is arguably not required for weakness of will. What if I maintain my resolution but make no attempt to act on it, due to the inclinations I hoped to resist?<sup>15</sup> This is weakness without revision. Or suppose I resolve to do something I know I shouldn't, despite my scruples, only to revise my intention, quite reasonably, in light of them. Can't this constitute weakness of will?<sup>16</sup>

We can set these questions aside. My interest is not the definition of weakness, or the empirical phenomena of addiction or temptation, but the mechanics of being strong-willed. Holton is struck by the phenomenology of active effort that accompanies the exercise of will in those who struggle with temptation or addiction. And he is impressed by empirical evidence of "ego depletion" in which the energy one has available to exert one's will is depleted by prior efforts. He concludes that belief-desire and belief-desire-intention models of intentional action are incomplete. Willpower is a further element of our psychology, neglected by these models, but essential to what happens in those who manifest strength or weakness of will (Holton 2009: 118-21).

I accept this conclusion, in a way. The active exertion of willpower is real and it is not reducible to shifts in belief, desire, or intention. But if we say no more than that, we leave the active nature of the will obscure. When we look at what's involved in this activity, what we find is not a single thing, "exerting one's will," but a wide variety of actions. In Walter Mischel's well-known "marshmallow experiment," cited by Holton, the children who hold out for a second marshmallow do so by reaffirming their intentions quietly to themselves, by covering their eyes so they cannot see the single marshmallow in front of them, by distracting themselves with songs or games (Holton 2009: 126-7). We exert willpower intentionally by doing things, some of them overt, others mental, like directing our attentional focus, our imaginings, our inner speech (Holton 2009: 123-5). These activities are no more reducible to belief, desire, and intention than any other intentional actions. We need to supplement the belief-desire-intention model to account for strength of will only in the sense that we need to

---

<sup>15</sup> This possibility is explored at length in Asarnow 2019.

<sup>16</sup> For this and other objections, see Yao 2017.

expand the range of intentional actions we consider to include activities by which we try to prevent our intention to  $\phi$  from being deflected by temptation when we have resolved to  $\phi$ . Such activities often accompany direct attempts to  $\phi$ , being performed alongside them. They are forms of self-control in that they are intentional actions by which we exercise control over ourselves, amplifying certain motivations and muting others. As Albritton writes: “controlling yourself is a *project*, possible or impossible, like controlling a dog” (Albritton 1985: 250).

The idea that strength of will involves the performance of ordinary intentional actions, not an irreducible exercise of willpower, dovetails neatly with the conditional analysis of freedom in its more sophisticated guise. If S would  $\phi$  if she resolved to  $\phi$ —that is, if she intended to  $\phi$  and intended not to let that intention be deflected by temptation—then she is able to  $\phi$ , even if, were she simply to intend to  $\phi$  without exercising self-control, she’d fail. She knows how to control the inclinations that would lead her not to  $\phi$ , perhaps by refocusing her attention, or distracting herself. There is a complex action,  $\psi$ , such that, if S intended to  $\psi$ , S would  $\phi$ , she knows that  $\psi$ -ing is a means by which to  $\phi$ , and she knows how to  $\psi$ ; so, by Ability, she can  $\phi$ .

Self-control is in part about knowing how to manage oneself—how to prevent one’s intentions from being deflected by temptation—in part about being so disposed that there are effective strategies for doing so: being responsive to self-directed inner speech, for instance. Where the tactics of self-control would be ineffective, and there are no other relevant means by which to  $\phi$ , S cannot  $\phi$  by the lights of Inability: for all  $\psi$ , if S intended to  $\psi$ , S would not  $\phi$ . This is the sort of compulsion that interests philosophers of action, the sort experienced by Albritton’s drinker and torture victim, the sort I imagined in Lehrer’s candyphobe. How prevalent it is in fact is an empirical question—made difficult in part by the challenge of saying what counts as a tactic of self-control. Acts of attention or inner speech are relatively uncontentious cases. But Mischel’s children used overt actions, too. Where to draw the line between mental acts that affect one’s motivations and actions that adjust one’s circumstance so that one’s motivations are less likely to deter one from acting as one intends? Often one reduces the intensity of a desire precisely by changing the circumstance that provokes it, as when

the child faced with a tempting marshmallow covers her eyes. Are her “inner resources” of self-control sufficient to withstand temptation? Or does this count as external help?

I don't see that much hangs on drawing these lines, though perhaps it does for the questions of culpability I have set aside. Even without more precision, this account of strength and weakness of will sheds light on a longstanding puzzle due to Gary Watson. In “Skepticism about Weakness of Will” (Watson 1977), he asked: in what sense is the weak-willed but not the compulsive agent able to resist temptation? After all, each of them intends to do something and fails to follow through because a conflicting desire is motivationally more powerful than their initial intention. Both act on their strongest motive; and it may be perfectly predictable, in each case, that they will give in. Watson concludes that the weak-willed agent is no more able to resist temptation than the one who is subject to compulsion, and proposes an alternative contrast between them, set by social expectations or norms.

But we can make distinctions of ability more fine-grained than the relative strength of intention and desire. When an agent is compulsive, she intends to  $\phi$ , and perhaps resolves to do so, intending to prevent her intention to  $\phi$  from being deflected by temptation. Yet no matter what tactics she decides upon, she will fail. Assuming that she has no other means by which to  $\phi$ , she is unable to  $\phi$ . Things are different with those who are merely weak-willed. For them, the tactics of self-control might work. This is not to say that there is any tactic that *would* work, although there may be. But there's a decision, joining the decision to  $\phi$ , that might have made a difference. If the agent does not know what this decision would have been, we may be inclined to attribute her failure to ignorance, not weakness of will. But the lines again seem vague. After all, the skill of self-control partly consists in knowledge-how; and the agent might fail to consider, or recognize, or take, what she knows to be potential means to self-control. The picture is appropriately blurry, but it would be wrong to say that no contrast of significance lies in the ability to do otherwise—which the compulsive agent, but not the weak-willed, lacks.

## V

At this point, one might hope for reflections on determinism and the ability to do otherwise, perhaps a frank defence of compatibilism about freedom and responsibility. One would be disappointed—except that, if I am right about freedom to act, what it requires is nothing more than that one’s actions counterfactually depend on one’s intentions.

Rather than pursue this thought into more substantive but treacherous terrain, I’ll retreat to the superficially deep. Albritton sets out to defend Descartes’ view that “the will is perfectly free, ‘so free in its nature that it cannot be constrained’” (Albritton 1985: 239). “I don’t exactly know what any of that means,” he admits, “but I don’t know how to doubt that we’ve got free will, either” (Albritton 1985: 243). Perhaps we can now say what this means and why it is both true and trivial.

According to Inability, “S cannot  $\varphi$ ” is true just in case, for all  $\psi$ , if S intended to  $\psi$ , S would not  $\varphi$ . When this condition is met, S is not free to  $\varphi$ ; she is unable to do so. Now, it is trivial that, if S intended to  $\varphi$ , S would intend to  $\varphi$ . But then it can’t be true that S cannot intend to  $\varphi$ , in the relevant sense. For there is always an action,  $\psi$ , such that, if S intended to  $\psi$ , S would intend to  $\varphi$ —namely  $\varphi$  itself. It follows that, whatever the value of “ $\varphi$ ” we can never truly say that S is not free to intend to  $\varphi$ , or that she is unable to do so, without changing the subject. If “can” is an ability modal, “S can’t intend to  $\varphi$ ” cannot possibly be true.<sup>17</sup>

Of course, for that very reason, we’ll want to interpret the sentence differently, as a possibility modal, not concerned with what S has the ability to do but with what could or might occur. On this reading, we can truly say, “S could never intend to  $\varphi$ .” And this fact may be relevant to responsibility or blame. But two points remain. First, we have no need to invoke such claims in order to explain the deficit of freedom in compulsion, which is accounted for by Inability. The compulsive cannot do otherwise in just the sense in which the person locked in chains can’t go for a walk: any intention to

---

<sup>17</sup> It doesn’t follow that S *can* intend to  $\varphi$ . Ability’s first condition is met, but its sufficient condition requires that S know that  $\varphi$ -ing is means by which to intend to  $\varphi$ , and knows how to  $\varphi$ —where the latter may not be true and the former is generally false.



that effect would fail. If that absolves them from responsibility, it does so without recourse to a distinctive “unfreedom of the will.” Second, claims about what it’s possible—or impossible—for someone to intend are not claims about freedom in anything like the sense in which we ascribe to agents the freedom or ability to act. If we focus on freedom in that sense, Albritton and Descartes are right: the will is free in its very nature—so free that it cannot be constrained.<sup>18</sup>

---

<sup>18</sup> Thanks to Karin Boxer, Sarah Buss, Justin Coates, Pamela Hieronymi, Claire Kirwan, Harvey Lederman, Matt Mandelkern, Derk Pereboom, Tim Schroeder, George Sher, Susan Wolf, and Katie Zhou, to the SHAPE reading group at MIT, and to audiences at Meaning and Morality: A Conference in Honor of Susan Wolf and at the MIT graduate alumni conference, Groundhog Day 2024.

## References

- Albritton, R. 1985. Freedom of Will and Freedom of Action. *Proceedings and Addresses of the American Philosophical Association* 59: 239-251.
- Albritton, R. 2011. On a Form of Skeptical Argument from Possibility. *Philosophical Issues* 21: 1-24.
- Asarnow, S. 2019. On Not Getting Out of Bed. *Philosophical Studies* 176: 1639-1666.
- Austin, J. L. 1956. Ifs and Cans. Reprinted in *Philosophical Papers*. Oxford: Oxford University Press, 1979: 205-232.
- Bhatt, R. 1999. *Covert Modality in Non-Finite Contexts*. PhD Thesis: University of Pennsylvania.
- Boylan, D. 2022. Does Success Entail Ability? *Noûs* 56: 570-601.
- Chisholm, R. 1964. J. L. Austin's *Philosophical Papers*. *Mind* 73: 1-26.
- Holton, R. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
- Hume, D. 1748. *An Enquiry Concerning Human Understanding*. T. Beauchamp, ed., Oxford: Oxford University Press, 1999.
- Lehrer, K. 1968. Cans without Ifs. *Analysis* 29: 29-32.
- Loets, A. and Zakkou, J. 2022. Agentive Duality Reconsidered. *Philosophical Studies* 179: 3771-3789.
- Mandelkern, M. ms. Chance, Ability, and Control.
- Mandelkern, M, Schultheis, G, and Boylan, D. 2017. Agentive Modals. *Philosophical Review* 126: 301-343
- Murdoch, I. 1970. *The Sovereignty of Good*. London: Routledge.
- Pickard, H. 2015, Psychopathology and the Ability to Do Otherwise. *Philosophy and Phenomenological Research* 90: 135-163.
- Setiya, K. 2016. *Practical Knowledge: Selected Essays*. Oxford: Oxford University Press.
- Setiya, K. 2022. What is Morality? *Philosophical Studies* 179: 1113-1133.
- Solomon, A. 1998. Anatomy of Melancholy. *New Yorker*, January 12: 46-58.
- Strawson, P. F. 1962. Freedom and Resentment. *Proceedings of the British Academy* 48: 187-211.
- Watson, G. 1977. Skepticism about Weakness of Will. Reprinted in *Agency and Answerability*. Oxford: Oxford University Press, 2004: 33-58.
- Watson, G. 2002. Volitional Necessities. Reprinted in *Agency and Answerability*. Oxford: Oxford University Press, 2004: 88-122.
- Wolf, S. 1990. *Freedom within Reason*. Oxford: Oxford University Press.
- Yao, V. 2017. Strong-Willed Akrasia. *Oxford Studies in Agency and Responsibility* 4: 6-27.