How Is Strength of Will Possible?[1]

Richard Holton

Weakness of will is traditionally identified with *akrasia*: weak-willed agents, on this view, are those who intentionally do other than that which they judge to be best. This gives rise to the puzzle of how such failure is possible: how can an agent intentionally perform an action whilst believing a better option is available?

Suppose, however, that one were unconvinced by the traditional identification of weakness of will with *akrasia*. Suppose one thought instead of weakness of will as failure to persist in one's resolutions. And, correspondingly, suppose one thought of strength of will as success in persisting with one's resolutions. Then the interesting question would no longer be how *weakness* of will is possible. It is all too easy to see how an earlier resolution could be overcome by the growth of a subsequent desire. Rather, the interesting question would be how *strength* of will is possible. How do agents succeed in persisting with their resolutions in the face of strong contrary inclinations?[2]

Elsewhere I have argued for an account of weakness of will and strength of will along these lines; I will summarize those ideas shortly.[3] Here my focus is on the interesting question that follows: on how strength of will is possible. My answer, in brief, is that we standardly achieve strength of will by exercising will-power. I mean this as more than a pleonasm. My claim is that will-power is a distinct faculty, the exercise of which causally explains our ability to stick to a resolution.

To get some idea of what a separate faculty of will-power might be, let us contrast this approach with the two alternatives that have been dominant in recent philosophical discussion (alternatives first):

I. The Humean Account (Belief/Desire Account)

This seeks to explain all intentional action in terms of the agent's beliefs and desires. Agents act on whichever of their desires are strongest.[4] An explanation

---

[1] This paper was given as a talk at the conference *The Will in Moral Psychology*, held in Edinburgh in July 2002. It is a précis of several chapters of an unfinished book, *Aspects of the Will*. Issues skated over here just might receive a more adequate treatment there. Thanks to the Edinburgh audience, the editors, and especially to Rae Langton and Alison McIntyre; also to the AHRB for a grant that gave me a year free of teaching, during which this was written.

[2] Kent Bach makes much the same point in his review of George Ainslie's *Picoeconomics*, *Philosophy and Phenomenological Research* 55 (1995) pp. 981–3. Note that in a case of weakness of will, the subsequent desires can affect the judgment about what is best; in such a case we need not have an instance of *akrasia*.

[3] 'Intention and Weakness of Will', *Journal of Philosophy* 96 (1999) pp. 241-62.

[4] More strictly we should factor in agents' beliefs about which of their desires can be realized: agents will be unlikely to act on their strongest desires in cases in which they think them probably unattainable, but think of other desires, nearly as strongly held, as readily attainable.

of how agents stick by their resolutions must show how they thereby act on their strongest desires. (In so far as resolutions are understood as mental states at all, they must be thus reducible to beliefs and desires.)

II. The Augmented Humean Account (Belief/Desire/Intention Account)

This holds that beliefs and desires won't do the job. Intentions, of which resolutions are a species, should be seen as a third mental kind, irreducible to the other two. This second account thus involves an *ontological* revision of the first. However, when it comes to the *mechanism* that explains strength of will, there is no fundamental change. For this account keeps the idea that it is the relative strength of the conative inputs that determines what the agent will do; it is just that these now consist not simply of desires, but of desires *and* intentions. If a resolution is stronger than any contrary desires, the agent will stick to it; if the contrary desires are stronger, then the agent will act on them instead.

III. The Will-Power Account

Like the second account, this keeps the idea that the basic mental states are beliefs, desires, and intentions. However, it differs radically in the mechanism by which the agent's action is determined. This is where the idea of will-power comes in. The claim is that the agent's decision is determined not just by the relative strength of the conative inputs, the desires and the intentions. Rather, there is a separate faculty of will-power which plays an independent contributory role. Agents whose will-power is strong can stick by their resolutions even in the face of strong contrary desires; agents whose will-power is weak readily abandon their resolutions.

My project here is to develop an account of this third sort. Part of the ambition is simply to show that there is the necessary conceptual space for it: the recent dominance of the other two accounts has tended to obscure the very possibility of taking will-power seriously. But, of course, I want to go further than that. I want to argue that there are considerations, both philosophical and psychological, that show its advantages over the others. I start with a brief summary of the conception of strength of will that makes these issues pressing.

Akrasia and Weakness of Will

Imagine someone who is convinced that all the arguments point the same way: he should give up meat. Yet suppose he does not. He is, therefore, *akratic*. But is he weak-willed? That, I contend, depends on other factors. Has he repeatedly vowed to give up, only to find himself succumbing time and again in the face of rare steaks and slow-cooked offal? Or does he unblushingly affirm whilst conceding practical inconsistency—that he has never had any intention of giving up, and never will? In the

---

To keep things manageable I ignore these complications here, though they are relevant to some of the motivational issues discussed below.

first case we surely would accuse him of weakness of will; in the second I think we would not. Although there is something very odd, indeed culpable, about his behaviour, it is not what we ordinarily think of as weakness of will.

If that is right, then central to the idea of weakness of will is an over-readiness to abandon one's resolutions. The weak-willed carnivore is the one who cannot maintain his resolve. How should that over-readiness be understood? Resolutions can be understood as a special kind of intention. Intentions themselves can play a number of roles; this is true whether or not we are reductivist about them. As Michael Bratman has shown, they serve to foreclose deliberation, and have an important function in enabling coordination.[5] But they can also serve to protect the outcome of earlier reasoning from later temptation. I might judge that it will be better for me to work this evening than to do any of a number of other tempting things. Currently, moreover, working is what I most want to do. However, I know that when this evening comes, besieged by temptations, I will more strongly want to do one of those other things. What can I do to protect my current judgement—that working is the thing to do—from the temptations that threaten to overwhelm it?

One thing I can do—indeed the most normal thing to do—is simply to *decide* now that I will work this evening. That is, I can form the intention to do so. But this is a special sort of intention. Its distinctive feature is that it is supposed to remain firm in the face of the contrary desires that I expect to have. This is what I take to be characteristic of a resolution. Resolutions are intentions part of whose function is to defeat contrary inclinations that I fear I might come to have.[6] Abandoning such intentions results in a special kind of failure. If new desires cause me to abandon my intention to eat in one restaurant in favour of an intention to eat in another, I open myself to the charge of being fickle. But when new desires cause to me to abandon a resolution, I open myself to the more serious charge of being weak-willed.

However, not every case of abandoning a resolution in the face of the desires it was supposed to defeat is a case of weakness of will. Sometimes I might form a resolution for a very trivial reason. I might, for instance, resolve to go without water for two days to see what it feels like. And sometimes the contrary desires will be far stronger than I had imagined: perhaps, after a day, the desire for water will be enormous. In such cases we might be reluctant to say that those who revise their resolutions are weak-willed. Indeed, the failing would lie with those who persist. They would exhibit an unreasonable inflexibility or stubbornness. So weakness of will involves, I think, a normative element. It is the *unreasonable* revision of a resolution in the face of the contrary desires (or inclinations more generally) that it was supposed to defeat.[7]

---

[5] M. Bratman, *Intention, Plans and Practical Reason* (Cambridge MA: Harvard University Press, 1987).

[6] In 'Intention and Weakness of Will' I called these 'contrary inclination defeating intentions'; 'resolutions' is somewhat easier to say.

[7] Alison McIntyre showed me that, for it to be a case of weakness of will, the desire that defeats the resolution must be of the kind that the resolution was designed to defeat—a point which I missed in my earlier piece. Note too that there can be other reasons for revising a resolution that do not bring the charge of weakness of will. I might, for instance, become convinced (even unreasonably convinced) that the premise on which the original resolution was based is false. Then revising the resolution (even unreasonably) would not exhibit

We might expect strength of will to be the contrary of weakness of will. I think that it is. The central feature of strength of will is the ability to maintain one's resolutions in the face of the very inclinations that they were designed to overcome. In addition I think that it too involves a normative element. To show strength of will is not to maintain one's resolutions come what may. That, as we have seen, can sometimes be mere stubbornness. Rather it is *reasonable* maintenance that is required.

In the last section we shall return to this normative issue. But for the bulk of this paper I shall be concerned with the descriptive issue of how strength of will is possible: how is it that we can maintain a resolution in the face of contrary desires. It is all too common to find that resolutions are overwhelmed by these desires; we need to know how it is that sometimes they are not.

Explaining Strength of Will in the Humean (Belief/Desire) Approach

My aim in this section is to describe what I take to be the most promising account of strength of will within the belief/desire framework, and to show that it fails. In so doing I hope to motivate the need for an alternative.

Let us start with a classic case of the need for strength of will. Suppose that you have a desire to give up smoking; that is, you prefer

A. I give up smoking for good soon

to

B. I don't give up smoking for good soon.[8]

However, you know that you will also strongly desire any particular cigarette that you are offered. "And why not?", you might think: "No single cigarette is going to do me much harm, yet the pleasure it will give will be great". So you know that, for each cigarette at the moment before smoking it, you prefer:

C. I don't resist this cigarette

to

D. I resist this cigarette.[9]

It is easy to see where this reasoning will lead. It seems that if you act on your strongest desires, you will always put off giving up until after the next cigarette; and so

weakness of will—provided that my conviction that the premise was false was not, in turn, simply a rationalization triggered by the desires that the resolution was supposed to defeat.

[8] I put aside complications that come from the possibility that smoking is addictive. If that worries you, substitute an innocuous example.

[9] I say that these are the preferences *at the moment just before smoking each cigarette*. If you were to think *at all times* that it was preferable to smoke each cigarette than not to, then these preferences would be simply inconsistent with your preference to give up: in wanting to give up you would simply have failed to sum your individual preferences properly. In the situation I have in mind, you avoid that kind of inconsistency since your preferences change in the proximity of a cigarette. In the terminology made famous by George Ainslie, you have hyperbolic discount curves. For Ainslie's account of such preferences, see his *Picoeconomics* (Cambridge: Cambridge University Press, 1992) and *Breakdown of Will* (Cambridge: Cambridge University Press, 2001). For a very clear presentation see H. Rachlin, *The Science of Self-Control* (Cambridge MA: Harvard University Press, 2000) Ch. 2.

you will never give up. This is true even if your desire to give up is greater than your desire for any particular cigarette (i.e. if you prefer A to C), since it seems that it is possible both to smoke any particular cigarette, and to give up in the near future.

It might appear then that the belief/desire account is in trouble right from the beginning. Given a pattern of desires that has this form—surely a very common one—it looks as though an agent who is motivated solely by desire will be unable to exercise strength of will. And that looks like a refutation of the belief/desire account, since surely agents with just this pattern of desires do sometimes display strength of will.

There are, however, two responses that the proponent of the belief/desire account can make. The first involves adding further desires; the second involves adding further beliefs. I take them in turn.

Adding a further desire

Although the belief/desire account makes do with just beliefs and desires, that does not mean that it can have no place for resolutions. They might be accepted as mental states, to be analyzed as a form of belief, or a form of desire, or some combination of the two. Alternatively they might be seen, not as mental states at all, but as something like illocutionary acts. The obvious model here is promising. On this second approach, when a person makes a resolution she makes something like a promise to herself. This will typically give rise to a mental state: to the belief that she has made the resolution. But that belief isn't itself the resolution. The resolution is the illocutionary act that the belief is about.

Either of these ways of accommodating resolutions, the reductive or the illocutionary, now provides for a possible way out of the problem. For suppose that you do not simply *desire* to give up smoking; in addition you *resolve* to do so, forming a resolution that bears some particular date for its implementation. And suppose that you have *a strong desire to be resolute*: a strong desire to stick to your resolutions. Then, when the date for implementing the resolution comes, provided that your desire to be resolute is stronger than your desire to smoke, you have a desire-driven way to give up. Unlike the desire to give up sometime soon, the desire to be resolute cannot be satisfied compatibly with taking the next cigarette after the resolution is to be implemented. The date on which you resolve to give up can be completely arbitrary, but it becomes significant because you choose it.[10]

Adding some further beliefs

An alternative tack works by adding further beliefs rather than further desires. Recall that the initial difficulty got going because at each point you thought that it was possible both to take the next cigarette, and to give up smoking some time soon. Suppose that you come to doubt that: suppose that at some point you come to believe that whether you give up smoking some time soon is dependent on whether you smoke

---

[10] For a discussion of an approach along these lines see J. H. Sobel, *Taking Chances: Essays on Rational Choice* (Cambridge: Cambridge University Press, 1994) pp. 249–50.

the next cigarette. Then you will be able to use your stronger desire to give up smoking soon (A) to overcome your desire to smoke the next cigarette (C). It is important to see what talk of 'dependent' here must mean. If the desire to give up smoking is to exert the requisite leverage, you must believe both

Effective: If I resist this next cigarette, I'll give up smoking for good

and

Necessary: If I don't resist this next cigarette, I won't give up smoking for good.

The names should make clear the functions of the beliefs, but let us spell them out nonetheless. If *Effective* is absent you will fail to think that resisting the next cigarette will have any effect in realizing your desire to give up soon, and so you will have no reason to resist. If *Necessary* is absent you can consistently think that you will be able both to smoke the next cigarette to still give up, so again your desire to give up will provide no reason for resisting this cigarette.[11]

Why should you come to believe both *Necessary* and *Effective*? *Effective* might be justified on simple inductive grounds. If you feared that you would be simply unable ever to resist a cigarette, then resisting one now will show that your fear was ungrounded. Perhaps too it will be underpinned by some kind of sunk-cost reasoning. The more suffering you have endured to resist cigarettes, the more likely you will be to be motivated to resist them in the future: what a waste of effort otherwise! We can accept that people are in fact motivated in this way whether or not we think, with most economists, that there is something irrational about it.

Necessary is harder to justify. Presumably in forming a resolution to stop smoking you will have chosen some particular point as the point at which to give up.[12] Then your conviction in *Necessary* might be underpinned by some kind of *now-or-never* thinking. You can accept that the point which you chose is arbitrary. Nevertheless, you can think that, having chosen this point, you must stick to it: if you break your resolution to give up smoking now, you will never be in a position to stick to a similar resolution at any point in the future.

Moreover, we can see how this reason for believing in *Necessary* might interact with the phenomenon of wanting to be resolute that was discussed above. You might think that a failure to stick to this resolution to give up smoking would adversely affect your ability to stick to any other resolutions that you might form, resolutions about things quite unconnected with smoking. And so, in so far as that is an ability that you

---

[11] This approach, though cast in a game theoretic context which views agents as collections of (potentially competing) time-slices, originates with George Ainslie in the works mentioned above. For criticism of the time-slice framework, see M. Bratman, 'Planning and Temptation' in *Faces of Intention* (Cambridge: Cambridge University Press, 1999) pp. 35–57. For a reworking of the approach without time-slices, resulting in a position similar to that presented here, see R. Nozick, *The Nature of Rationality* (Princeton: Princeton University Press, 1993) pp. 14–21, and A. Mele, 'Addiction and Self-Control', *Behavior and Philosophy* 24 (1996) pp. 99–117.

[12] If you didn't, your resolution is unlikely to succeed; there is good evidence that resolutions without 'implementation intentions' are far less effective. See P. Gollwitzer, 'The Volitional Benefits of Planning' in P. Gollwitzer and J. Bargh (eds.), *The Psychology of Action* (New York: The Guildford Press, 1996) pp. 287–312.

strongly want to keep, you have a further motivation for sticking to this particular resolution.[13]

Problems

So we have two attempts to explain strength of will within the belief/desire framework. Both involve ideas that have some plausibility. Yet neither, I think, will work as a complete account. For a start, both are vulnerable to serious problems of detail. [14] These, however, won't be my concern here. More fundamentally, both completely misrepresent the phenomenology of the exercise of strength of will.

The central point is this. If these accounts were right, then sticking to a resolution would consist in the triumph of one desire (the stronger) over another. But that isn't what it feels like. It typically feels as though there is a struggle: that one maintains one's resolution by dint of effort in the face of the contrary desire. Perhaps not every case of maintaining strength of will is like that (we shall mention some that are not). But by-and-large, maintaining strength of will requires effort.

Moreover the empirical evidence bears this out. The most straightforward comes from simple measures of the physical arousal to which the exercise of will-power gives rise. Ask agents to regulate themselves in ways which involve acting against contrary inclinations—to regulate their emotions for instance, the expression of their emotions, their attention or their thoughts—and they will show the standard signs of physiological arousal that accompany effort: increased blood pressure and pulse, changed skin conductance etc.[15]

Of course it is true that cases only involving desires can give rise to a feeling of struggle: consider what it can be like trying choose an option from a wide and attractive menu, even when there are no resolutions to which one is trying to keep. Here too we might speak of the *effort* of choosing. We might well think that this in itself shows that there is something wrong with the Humean picture: that in general it leaves insufficient space for the role of the active agent in making choices. However, my concern here is with the particular case of choice constrained by resolution. If the Humean account were right, we would expect the phenomenology of the effort of choosing to be the same as the phenomenology of the effort of maintaining a

---

[13] The parallel again is with promising, understood in a broadly Humean way: resolutions are devices that enable you to stake your general reputation on each individual case.

[14] The *further desire* approach seems to involve attributing to the strong-willed agent a desire for resoluteness that approaches a fetish. The *further belief* approach faces difficulties in establishing that a reasonable agent would believe both *Necessary* and *Effective*. Why not think, for instance, that *Effective* would be undermined by the inductively sustained belief that, at least for the first few weeks, resolutions become harder to maintain as time goes on? It is easy enough to refuse the first cigarette; the difficult thing is to keep on refusing. Similarly, why wouldn't *Necessary* be undermined by the knowledge that many agents only give up smoking after several attempts to do so? In their cases the failure of one resolution didn't entail the failure of all subsequent ones.

[15] M. Muraven, D. Tice and R. Baumeister, 'Self-Control as a Limited Resource: Regulatory Depletion Patterns', *Journal of Personality and Social Psychology* 74 (1998) pp. 774–789 includes a survey of this at pp. 774–5.

resolution that one has already chosen. But they are not. And the kind of regret that one can feel for an option not chosen is very different from the kind of regret—to say nothing of shame or guilt—that one feels when one abandons a resolution.[16]

Intentions and Will-Power

How can we make sense of this idea of struggle? A first move is to distinguish resolutions from desires, for only then can we make sense of the idea of struggle involved in sticking with a resolution rather than bending to a desire. Resolutions, I have suggested, can be seen as a particular kind of intention. Having rejected the Humean account, we can follow Michael Bratman and others in treating intentions as mental states that are distinct from (i.e. not reducible to) beliefs and desires.[17] Nevertheless, like desires, they are motivating states: an intention can move one to action. Intentions can thus work to preserve the motivational power of earlier desires: a desire can give rise to an intention, and this intention can result in subsequent action even when the desire is no longer present. Indeed an intention can result in subsequent action even when there are, by that time, contrary desires present. That, I suggested earlier, is precisely the role of resolutions. Resolutions are contrary inclination defeating intentions: intentions formed by the agent with the very role of defeating any contrary inclinations that might emerge.

Let us spell out some of the respects in which an approach that treats resolutions in this way departs from the Humean theory of motivation. It is not the case that to be motivated to act, an agent requires a belief and a desire. Nor is it true that agents will always act on their strongest desires. For an intention can serve as a motivation even when the desires that gave rise to it have been lost. Moreover, this intention can overcome the desires that are present at the time of action.

Once we introduce intentions in this way, how should we fill out the account? One possibility is to preserve something of the spirit of the Humean account. We might simply increase the class of motivating attitudes to encompass intentions as well as desires. Then, rather than saying that agents will act to satisfy their strongest *desire*, we might say that they will act to satisfy whichever is the strongest of their *desires and intentions*. Thus agents' actions will be determined by their beliefs, desires and intentions. This takes us to the second of the accounts of strength of will that was mentioned at the outset: the Augmented Humean account.

Alternatively we could move further still, to the account I shall defend: the will-power account which acknowledges beliefs, desires and intentions but adds a distinct faculty of will-power as well. How does this change things? One obvious difference is that here the strength of the agent's desires and intentions is not the only determinant of what she will do. We also need to add the strength of her will-power as a separate factor. Putting things in these terms can, however, be misleading, for it suggests a picture in which will-power is simply a third input in the process that determines the

---

[16] Although I say this so confidently, there is some empirical evidence that the effort of choosing has the same kinds of effect as the effort of maintaining a resolution. This might suggest that the same faculty is used in the two cases. I discuss this below, n.35.

[17] Bratman, *Intention, Plans and Practical Reason*.

agent's actions, a process on which the agent will seem like a spectator. I want rather to defend a picture on which will-power is faculty which the agent actively employs. The extent to which this can be achieved, will emerge, I hope, in what follows.

What, then, are the grounds for preferring the will-power account over the apparently simpler Augmented Humean alternative? My main contention is simply that it is better supported by the empirical evidence, both from ordinary common sense observation, and from psychology. Indeed, the psychological literature does not just provide evidence for the *existence* of a distinct faculty of will-power which works to block reconsideration of past resolutions; it also provides some quite detailed evidence about the *nature* of that faculty. Roughly, it seems that will-power works very much like a muscle, something that it takes effort to employ, that tires in the short run, but that can be built up in the long run.

I shall present some of the psychological evidence shortly. But to see its relevance, first let us return to the commonplace observation that we used in rejecting the simple belief/desire approach: the observation that exercising will-power takes effort. Sticking by one's resolutions is hard work. This seems to count against the Augmented Humean account too. 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 which requires effort.

My suggestion is that effort is needed because one is actively employing one's faculty of will-power. What exactly does the effort consist in? It cannot be straightforward physical effort, since it is present whether the resolution is to perform an action—like starting on an exercise regime—or to refrain from performing an action—like giving up smoking. However much the desire might seem to drag one towards it, we cannot think that the effort of resisting literally consists in pulling back muscles that are straining for the cigarette. Rather, the effort involved has to be a kind of mental effort. It is the mental effort of maintaining one's resolutions; that is, of refusing to revise them. And my suggestion here is that one achieves this primarily by refusing to *reconsider* one's resolutions. On this picture then, the effort involved in employing will-power is the effort involved in refusing to reconsider one's resolutions; and the faculty of will-power is the faculty that enables one to achieve this.

Before discussing the relevant empirical literature, we need to get a little clearer on the distinction between revision and reconsideration that is invoked here. We will also need to get clear on a further distinction between reconsideration and the simple rehearsal or reminder of the reasons for which one is acting.

Revision, Reconsideration, and Rehearsal

To revise one's intentions is to change them; that much is clear. Obviously reconsiderations differ in that they do not have to result in change. But I suggest that the full-blown reconsideration of a resolution does involve *suspension* of that resolution. To fully reconsider a resolution is to open oneself to the possibility of revising it if the considerations come out a certain way; and that is to withdraw one's current commitment to it. Of course, one might say that the resolution remains in place pending the outcome of the revision. But such a claim does not carry conviction. For much of the point of a resolution, as with any intention, is that it is a fixed point

around which other actions—one's own and those of others—can be coordinated. To reconsider an intention is exactly to remove that status from it.[18]

Although to suspend a resolution is not, *ipso facto,* to revise it, when temptation is great it is hard to keep the two separate. Very often the force of one's desire will *akratically* overwhelm one's judgment; or it will corrupt that judgment, so that whilst what one does is what one judges best, this is not what one would have judged best in a cooler moment; or it will move one so quickly to abandon one's resolution that one will never even carry out the judgement as to whether this is the best thing to do. Suspending a resolution can be like removing the bolts on a sluice: although one only meant to feel the force of the water, once the bolts are gone, there is no way of holding it back.

At the other extreme from full-blown reconsideration is the state of not thinking about one's resolutions at all: form them and then act on them, without so much as contemplating them or the possibility of acting otherwise. Perhaps this is the idea that we have the very strong-willed individual who, as we might say, is never really tempted by the alternatives. It might seem then that this is what we should aim for with our resolutions. In fact in typical cases it would not work.

This kind of unthinking pattern best describes those actions that are automatic. Force yourself to get up at six every morning to go for a run, and after a while it will probably become automatic. The alarm clock will go, you will get out of bed, put on your running kit, and get outside without really giving thought to what you are doing. Much recent work in social psychology has shown just how widespread automatic behaviour is; there is even evidence to suggest that it involves quite different parts of the brain to those which are involved in volitional behaviour.[19] But the point at which an action becomes automatic is really the point at which will-power is no longer needed. There is good reason for this. At least to begin with, a resolution is typically a resolution to reform one's behaviour into paths which are not automatic. Indeed standardly the automatic behaviour is exactly the behaviour that one has resolved to stop—lighting up a cigarette for instance. If one is to be successful in resisting having a cigarette, and if cigarettes are around, one must constantly monitor whether or not one has picked one up; and one can hardly do that without thinking about cigarettes, and the possibility of smoking them. Successful resolutions cannot work unthinkingly.[20]

So to maintain a resolution like giving up smoking we need something in between full-blown reconsideration and unthinking action. Most resolutions are, I suspect, like this. What we need is a state that involves awareness of the resolution, and perhaps of the considerations for which it is held, but which doesn't involve reconsideration. The

---

[18] *Ibid.* p. 62.

[19] On the former point, see, for instance, J. Bargh and T. Chartrand, 'The Unbearable Automaticity of Being', *American Psychologist* 54 (1999) pp. 462–79. On the latter see M. Jahanshahi and C. Frith, 'Willed Action and its Impairments', *Cognitive Neuropsychology* 15 (1998) pp. 483–533 (my thanks to Michael Scott for making me aware of the neuropsychological literature).

[20] The need for self-monitoring is central to Carver and Scheier's feedback account. See C. Carver and M. Scheier, *On the Self-Regulation of Behaviour* (Cambridge: Cambridge University Press, 1998).

crucial factor here is that the resolution is not suspended. To remind oneself of one's resolutions is not, by itself, to bring them into question. (One can inspect the sluice bolts without removing them.) It is important that it is not suspended. For, as we have seen, once a resolution is suspended, it will all too easily be revised. We thus need a state of awareness that falls short of suspension: what I shall call *rehearsal*.

I speak as though the contrast between reconsideration and rehearsal is a sharp one. In fact, of course, there will be many states in between: what I have marked out are the extremes of a continuum. Moreover, very often mere rehearsal will lead one into reconsideration. This is unsurprising when one's rehearsal leads one to dwell on the benefits to be gained by yielding to temptation; but empirical work shows that the same effect will often come even when one's focus is on the benefits to be gained by holding out.[21]

Can we resist the slide from rehearsal to reconsideration by dint of mental effort? It might seem that this would require an ability to repress thought. The difficulty with such advice is that it is very hard to control one's thoughts directly. Indeed, the effort is typically counterproductive: attempting to repress a thought leads one to dwell on it all the more.[22] But need it be that mental control involves such direct repression?

In seeing the possibilities it is useful to look to the advice given by those professionally concerned with the business of resisting temptation. Here is a representative passage from Ignatius of Loyola, founder of the Jesuits:

> There are two ways of gaining merit when an evil thought comes from outside: the first ... I resist it promptly and it is overcome; the second I resist it, it recurs again and again and I keep on resisting until the thought goes away defeated ... One sins venially when the same thought of committing a mortal sin comes and one gives ear to it, dwelling on it a little or taking some sensual enjoyment from it, or when there is some negligence in rejecting this thought.[23]

Quite what does 'resisting' a thought amount to? It does not seem that Ignatius is calling for out-right thought suppression. Rather he talks of the risks of dwelling on a

---

[21] Focusing on the benefits to be gained from resisting temptation tends to make agents more likely to succumb, since, under the influence of the temptation, those benefits are judged less valuable. See R. Karniol and D. Miller, 'Why not wait? A cognitive model of self-imposed delay termination', *Journal of Personality and Social Psychology* 45 (1983) pp. 935–42. Note that this effect only occurs when the rewards of holding out are broadly comparable to the rewards of yielding. When the rewards of holding out are judged to be much larger, focusing on them seems to strengthen resolve. Similar findings appear in the work of Walter Mischel discussed below, although here things are complicated by the fact that the reward for resisting temptation was just more of what one would have got had one succumbed. Hence it is hard to distinguish thought about the reward for wanting from thoughts about the reward for succumbing.

[22] D. Wegner, *White Bears and Other Unwanted Thoughts* (New York: Viking Press, 1989); J. Uleman and J. Bargh (eds.), *Unintended Thought* (New York: The Guildford Press, 1989). Again, it has been suggested that this is connected with the idea of self-monitoring: in order to be sure that one is not thinking about something one needs to monitor that one is not: D. Wegner, 'Ironic Processes of Mental Control' *Psychological Review* 101 (1994) 34–52.

[23] Spiritual Exercises, ¶33–5, in *Personal Writings* (Harmondsworth: Penguin, 1996) p. 291. My thanks to Annamaria Schiaparelli for her advice to look at Ignatius.

thought, or of taking some sensual enjoyment from it. The idea seems to be, not that we can keep certain thoughts out entirely, but that we can avoid focusing on them and developing them. Here it does seem far more plausible that we have some control.

I know of no studies on this, but some light might be shed by considering some parallel cases, even if the parallel is far from perfect. Suppose I ask you not to think of the number two. That is almost impossible, and the very effort of monitoring what you are doing makes your failure all the more certain. But suppose I ask you not to multiply 345 by 27. Unless you are extraordinarily good at mental arithmetic, so that the answer simply jumps out at you, you won't find my request hard to comply with at all. Nor will your monitoring of what you are doing undermine your compliance. Similarly, suppose I ask you not to think through, in detail, the route that you take from home to work. You might not be able to resist imagining the starting point; but I suspect, unless you live very close to work, that you will be able to stop yourself somewhere down the track. The point seems to be that there are quite a few steps needed to perform a long multiplication or to imaginatively trace one's route, steps that have to be taken in a particular order, and one is able to exercise some control over such a process.

I suggest that things are typically similar with the thoughts involved in the revision of resolutions. It might be impossible to control whether we entertain the thought of having a cigarette. But it might be possible to control whether or not we go through the procedure that is involved in revising one's resolution not to. This also seems to be the kind of thing that Ignatius has in mind. The sin does not consist in having the evil thought that 'comes from outside'; Ignatius accepted that that is inevitable. The sin comes with what I do with it.

Evidence for Will-Power

My suggestion, then, is that whilst rehearsal is needed for maintaining a resolution, reconsideration should be avoided. Let us turn to the empirical evidence for this. I start with a discussion of the developmental evidence that suggests that resolutions really do work to block reconsideration in the way I have said. I then turn to the other considerations which show that abiding by a resolution does involve the exercise of a distinct faculty of will-power, a faculty which the agent actively employs.

Developmental evidence

Walter Mischel and his colleagues tested children on their ability to delay gratification to achieve greater reward.[24] For instance, they are told that they will receive one cookie if they ring a bell, which they are free to do at any time; but that they will get two if they refrain from ringing the bell until an adult comes in. They found that ability to wait comes in around the age of four or five. By the age of six almost all children have it, though to markedly different degrees. Strong self-control is a very good predictor of later success in a wide range of academic and social skills.

---

[24] For a summary of a large body of work see Mischel 'From Good Intentions to Willpower' in Gollwitzer and Bargh, *The Psychology of Action* pp. 197–218.

What are the strategies that children used? Mischel initially expected them to do better by being reminded of the rewards of waiting. In fact, however, those who could see the reward for waiting did far worse than those who could not. Those who could see the reward for not waiting did equally badly. Mischel's account is illuminating and entertaining enough to be worth quoting at length:

> Some of the most effective strategies that the children used were surprisingly simple. From time to time they reaffirmed their intentions quietly ("I'm waiting for the two cookies") and occasionally they reiterated the choice contingency aloud ("if I ring the bell I'll get this one, but if I wait I'll get those"). But mostly these 4-year-olds seemed able to wait for long periods by converting the frustrating waiting situation into a more tolerable non-waiting one, thus making the difficult task easier for themselves. They appeared to achieve this by purposely creating elaborate self-distraction. Instead of fixing their attention and thoughts on the rewards, as initially theorizing had predicted, they seemed to avoid thinking about them entirely. Some put their hands over their eyes, rested their heads on their arms, and invented other similar techniques for averting their gaze most of the time, occasionally seeming to remind themselves with a quick glance. Some talked quietly to themselves or even sang ("This is such a pretty day, hooray"); others made faces, picked their noses, made up games with their hands and feet, and even tried to doze off while continuing to wait. One of the most successful "delayers" actually managed to nap during the delay time.[25]

Here the children do seem to conform to the model I have proposed. They sometimes rehearse their resolution, and the reasons for having it (though in this case there is little benefit from so doing, since there is little need for self-monitoring). Seeing the cookies—whether the one to be gained by ringing the bell, or the two to be gained by waiting—radically undermined the children's ability to wait. It seems that this undermines resolve because it provokes reconsideration.[26] In a further series of experiments Mischel found that being able to see the rewards did not undermine the resolution if the children were encouraged to see them as in some way unreal. A plausible explanation is that thinking of the rewards in this way does not encourage reconsideration since they are not being thought of as the objects (or at least, not as the objects with the salient 'hot' properties) about which the resolution was made.[27]

Mischel's findings do, however, raise one question. Consider the children who had very effective strategies for distracting themselves. Considered in a behaviouristic way these might be thought of as those exercising the greatest will-power, since they are the ones who are most successful at resisting temptation. This is how Mischel describes them. But in another sense we might think of them as the children who have least need for will-power; after all, these are the ones who are putting in little effort since their strategies are so effective (think of the child who took a nap). I suspect that our ordinary talk of will-power is ambiguous here. In this it is no different from our talk of many other virtues. Are they brave who distract themselves in the face of danger? Or need they fight to overcome their overcome their fear? I doubt that our ordinary usage provides an answer. Similarly, I doubt that our ordinary usage dictates whether to be

---

[25] *Ibid.* p. 202.

[26] *Ibid.* pp. 201–2. This finding is corroborated by the Karniol and Miller study cited above.

[27] *Ibid.* pp. 203–7.

exercising will-power an agent has to be involved in an effortful struggle. However, to avoid confusion, I will legislate. I will limit talk of will-power to situations of effortful refusal to reconsider a resolution. In some cases then—such as the case of the automatic early morning runner considered earlier—agents achieve strength of will without recourse to will-power at all. It is unlikely that any of the children in Mischel's experiments were in quite that situation. They all had to employ will-power initially. But some had no need to go on employing will-power, exactly because their initial employment had been so effective.

Evidence for will-power as a separate faculty

The considerations marshalled so far support the idea that one exercises will-power by refusing to reconsider an intention. This in turn suggests that the Augmented Humean account is going to be inadequate, for one's ability to refuse to reconsider is not going to be determined just by the strength of one's desires and intentions. One does not acquire a practical ability just by wanting it. But we might wonder whether this does much to show that will-power is a separate faculty. Here we need to turn to some evidence from social psychology.[28]

Consider first the fact that the ability to abide by a resolution is affected by features that do not themselves seem to be desires or resolutions. Reformed alcoholics are far more likely to relapse if they are depressed, or anxious, or tired. [29] Moreover states such as these affect one's ability to abide by *all* of one's resolutions: resolutions not to drink, not to smoke, to eat well, to exercise, to work hard, not to watch daytime television, or whatever. Now of course it is possible to explain this by saying that these states (depression, anxiety, fatigue etc.) systematically strengthen all of one's desires to drink, smoke, eat, etc., or weaken all of one's resolutions not to; but it is surely a more economical explanation to say that they affect one's ability to act in line with one's resolutions.[30] For why else would there be such systematic effects?

Consider next the remarkable empirical literature on what is known as 'ego depletion'. It appears that will-power comes in limited amounts that can be used up: controlling oneself to eat radishes rather than the available chocolates in one

[28] For an excellent general survey of the relevant literature here see M. Muraven and R. Baumeister, 'Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?', *Psychological Bulletin* 126 (2000) pp. 247–59. Talk of *self-control* here, and elsewhere in the psychological literature, is, I think, roughly equivalent to my talk of *strength of will*. I would rather use *self-control* to describe the related but distinct phenomenon which is the contrary of *akrasia*: on this usage one lacks self-control if one does other than that which one judges best, even if one does not thereby violate one's resolve (and hence is not weak-willed).

[29] R. Baumeister, T. Heatherton and D. Tice, *Losing Control* (San Diego: Academic Press, 1994) pp. 15ff. The same is true of those who are dieting (*ibid.* pp. 184ff.), or trying to give up smoking (*ibid.* pp. 212ff.) or taking drugs (Muraven and Baumeister, 'Self-Regulation and Depletion of Limited Resources' p. 250).

[30] Moreover, whilst bad moods make dieters want to eat more, they tend to have the opposite effect on those who are not on a diet. So it seems that it is the resolution being affected, not the desire. See Muraven and Baumeister 'Self-Regulation and Depletion of Limited Resources' p. 251.

experiment makes one less likely to persist in trying to solve puzzles in the next[31]; suppressing one's emotional responses to a film makes one less likely to persist, later on, in holding squeezed a handgrip exerciser.[32] Again it is possible to think that what happens here is that the strength of people's resolutions are affected: that maintaining one's resolution to suppress one's emotional responses weakens one's resolution to persist with handgrip exercises. But why should there be effects on such disparate resolutions? And why do some activities (those that involve will-power to act in the face of inclinations to the contrary) bring about these effects, whilst others (such as doing mathematical problems) do not?[33] And why do dieters whose will-power has been tested react by subsequently eating more, whilst non-dieters do not?[34] A much better explanation is that one's action is determined not simply by the strength of one's desires and one's resolutions, but also by one's will-power; and that it is this component that is being affected by repeated exercise.[35]

A final piece of evidence is that one can apparently develop one's faculty of will-power by repeated exercise. Again, the idea that one becomes virtuous by behaving virtuously is a commonplace one, stressed by Aristotle and by many who have followed him: "From holding back from pleasures we become moderate, and also when we become moderate we are most capable of holding back from them"..[36] Some recent

---

[31] R. Baumeister, E. Bratslavsky, M. Muraven and D. Tice, 'Ego-depletion: Is the Active Self a Limited Resource?', *Journal of Personality and Social Psychology* 74 (1998) pp. 1252–65. The puzzles were in fact insoluble.

[32] Muraven, Tice and Baumeister, 'Self-Control as a Limited Resource'.

[33] *Ibid* pp. 781–2.

[34] K. Vohs and T. Heatherton, 'Self-Regulatory Failure: A Resource-Depletion Approach', *Psychological Science* 11 (2000) pp. 249–254. This finding parallels the finding of the effect on dieters of bad moods mentioned above.

[35] A note of caution is needed here. Whilst these experiments do suggest that there is a faculty of will-power, they do not give any evidence for my conjecture that it works by blocking reconsideration. Indeed, one of the experiments might be thought to raise a problem for this conjecture, since it suggests that *all* choices give rise to ego depletion (Baumeister, Bratslavsky, Muraven and Tice, 'Ego Depletion', Experiment 2, pp. 1256–8.) I cannot give a full description and discussion of this here: readers who are interested should consult the original article. However, two brief comments (i) I am unconvinced that this experiment does lend support to the conclusion that *all* choice gives rise to ego depletion, since the subjects make their choices under heavy moral pressure from the experimenters. It might well be this aspect (which is similar in kind to the pressure exerted by a resolution) rather than the choice *per se*, which gives rise to the ego depletion; (ii) even if it does turn out that all decisions involve ego depletion, I doubt that this would provide evidence against the hypothesis that will-power works by means of a refusal to reconsider. Rather, what it shows is that all decisions also involve a refusal to reconsider. But isn't that just what we should expect? On Bratman's account, intentions involve the foreclosure of deliberation. What is it to foreclose deliberation other than to refuse to reconsider? However, if this were true my earlier argument against the Humean theory would need revision. The Humean theory would still stand refuted, but not on the grounds that it makes exercises of will-power too similar to instances of choice.

[36] *NE* 1104a34. Aristotle is here talking about how we develop the excellences. He does not explicitly say the same about the development of self-control though. He does say that lack of self-control can be cured, but he doesn't say how.

research suggests that this might be right: subjects who undergo a regime of self-regulatory exercises—working on improving their posture for instance—show markedly less tendency to suffer ego-depletion.[37]

Once we think this way, a host more explanations become available. We have looked so far at intrapersonal differences: why do we sometimes stick by our resolutions, and sometimes not? But parallel explanations apply in interpersonal explanations: why do some people stick by their resolutions when others don't? It could be because their resolutions are stronger, or because the desires that they must overcome are weaker. Alternatively, it could be that their will-power is stronger: that having formed a resolution not to be moved by certain desires, they are better at acting in accordance with it, and at turning the corresponding intentions into action.

The approach employed here is a very general one. It has been a central feature of the cognitivist revolution that mental explanations, like explanations elsewhere, can be improved by positing further internal processes. Of course it is always possible to insist that wherever there is an intentional action, it is determined solely by strength of our pro-attitudes, whether these are understood just as desires, or as desires or intentions; and then to read off strength of those attitudes accordingly. But such an approach is not only untrue to our experience of sticking to a resolution; it also gives us, as the experiments I have cited show, inferior explanation of the behaviour that agents exhibit.

MOTIVATION AND WILL-POWER

I have argued that there is faculty of willpower—something like a muscle—and that, when desires and resolutions clash, we can succeed in sticking to our resolutions by employing this faculty. Moreover, employing the faculty is hard work: it requires effort on the part of the agent. What implications does this have for our explanations of why people do and do not stick to their intentions?

Obviously one class of explanations becomes immediately available. If agents lack will-power, then they will not be able to stick to their intentions in the face of contrary desires. This might happen as a result of never having acquired a faculty of will-power (as in the case of a young child) or from having lost it temporarily (from stress or

---

[37] M. Muraven, R. Baumeister and D. Tice, 'Longitudinal Improvement of Self-Regulation Through Practice: Building Self-Control Strength Through Repeated Exercise', *The Journal of Social Psychology* 139 (1999) pp. 446–57. Note that there was no effect shown on the *power* of the subjects' wills; only on their *stamina*, i.e. the degree to which they became fatigued. One thing that this research doesn't establish is whether the effect really comes from strengthening the faculty of will-power or from increasing the subjects' confidence that their resolutions will be effective. Indeed the further finding that attempts to implement resolutions in which it hard to succeed (control of mood) don't have the same effect on will-power might be explained that by the hypothesis that we are observing a self-efficacy effect (In general *self-efficacy*—one's confidence in one's degree of control—is extremely import in explaining one's behaviour. For a general overview see A. Bandura, 'Exercise of Personal Agency Through the Self-Efficacy Mechanism' in R. Schwarzer (ed.), *Self-Efficacy* (Bristol PA: Taylor and Francis, 1992) pp. 3–38). As the authors accept, we need more research here before any firm conclusions can be drawn.

fatigue or whatever) or perhaps even permanently (from damage to the pre-frontal cortex).

So some cases of failure to stick by a resolution will be explained by the absence of sufficient will-power. Will all explanations be of this form? That would require that agents always stick by their resolutions when they possess the will-power to do so: that the presence or absence of sufficient will-power is the only factor. Yet that is most implausible. If will-power is a faculty which agents actively employ, then it should be something that they can fail to employ. Surely sometimes people have the will-power to stick by a resolution and yet decide not to do so. I have resolved not to have wine with my dinner, and I know full well that I could resist, but I decide to have it anyway: the wine appeals very strongly, and I am not much moved by the need to keep a clear head afterwards. Such cases are very common. Indeed, even in cases where will-power is depleted by stress or prior demand or whatever, it seems likely that I will typically abandon the effort to stick by the resolution before my will-power gives way completely. It is not that I could no longer resist; it is that the effort becomes too great and I give up the fight.

Here again the analogy of the muscle, and of muscular fatigue, is helpful. Recall the subjects in the ego depletion experiments who were asked to hold squeezed a handgrip exerciser. We can easily imagine what it was like for them. The first few seconds were easy. Then, as the muscles got tired, they got more difficult. The hand started to ache, and the ache became more and more pressing, until the subject let go. We can imagine someone going on until the muscles literally could work no more. That is the kind of behaviour one sometimes sees in sporting competitions: grimacing, the competitor keeps on with the pull-ups, arms quivering uncontrollably, until, finally, the muscles give way. In such cases there is, quite literally, nothing more that the person could have done. In contrast, I doubt that any of the subjects with the handgrip exerciser pushed themselves so far. They got to a point where there said, perhaps even to themselves, they could go no further; but offered a large financial incentive, I suspect that they would have managed a few more seconds.

It is the ordinary handgrip subject rather than the competitive athlete who provides the better model for the typical defeat of will-power. Normally one does not find oneself literally powerless to resist a desire; rather, one *decides* to give in to it, since resistance is so hard (often, at the same time, convincing oneself that there is no good reason to resist.) A subject whose will is weakened by fatigue or prior demand simply finds the effort of resistance greater, and so typically gives up earlier. It is as though the handgrip subject started with an already tired hand. Of course in this case, fatigue of the hand muscles accompanied exhaustion of will-power (though the two process didn't quite walk in step: those whose wills had been earlier depleted presumably didn't start with fatigued hand muscles). In other cases there will be no concomitant muscular fatigue. The effort of resisting a cigarette is not literally a muscular effort; but it is no less real for that.

In determining whether agents will stick with their resolutions we need then to factor in not just their immediately relevant beliefs, desires, and intentions, and the strength of their faculty of will-power, but also their motivation to employ that faculty. And this motivation will in turn be cashed out in terms of further beliefs, desires, and intentions. Does this mean that we are back with a Humean model, or at

least with an Augmented Humean model, in which actions are determined by the strength of the beliefs, desires and intentions? It does not. An analogy might be helpful. If you want to know how fast I can run a mile on a given occasion, you'll certainly need to know about my beliefs and desires. Have I been offered some reward for running it fast? Will an embarrassingly poor time be publicized? But you will also need to know about the state of my body: is it any good for middle-distance running? It is the same for sticking to resolutions. If you want to know whether someone will stick to a given resolution you'll need to know about their beliefs and desires, including their desires with respect to the content of that resolution and with respect to maintaining resolutions in general. But you'll also need to know about their will-power: how strong it is, how much it has been depleted, and so on.

At this point, however, proponents of the Humean model might object. Isn't saying that one must be motivated to use one's will-power tantamount to saying that one must desire to use it above all? And isn't that just introducing once again the *further desire* account within the belief/desire model? I think not. What this challenge fails to recognize is the radical difference between intentions and desires. Intentions motivate directly: to act on an intention one doesn't need a further desire to act on that intention. Similarly, in the special case of resolutions, to act on a resolution one doesn't need a desire to act on that resolution, or on resolutions in general. For many agents in many cases, a resolution will simply work on its own; the agent's desires will be irrelevant. However, agents will be tempted to revise resolutions when acting upon them requires a large amount of effort. Whether or not they will do so will depend on, amongst other things, the strength of their desire to maintain those resolutions in particular, and the strength of their desire to maintain their resoluteness in general. But even here, to be effective, these desires need not be the strongest. If the agent's will-power is sufficiently strong, a weak desire to be resolute might be all that is needed to keep it in place when it wavers in the face of a strong contrary desire.

It is here, I think, that the true importance of the considerations raised in attempting to defend the belief/desire account come in; and this explains their plausibility. A desire to be resolute does indeed help an agent to be resolute, but it needn't be the overwhelming desire that the further-desire account held it to be. Similarly, belief in *Necessary* and *Effective* is highly relevant to whether agents will persist in their resolutions. An agent who has no confidence at all in *Effective*—an agent who fails to believe that if she refuses this cigarette then she will refuse others—will have little motivation to persist in her resolution. So even if she has the necessary will-power, it will not be used. Of course, an agent who knows that she has the necessary will-power will be far more likely to believe *Effective*, and so an absence of belief in *Effective* is likely to mark those who lack it. But it need not: will-power and self-knowledge need not go together.

The situation for *Necessary* is more plausibly the other way round. It is not that to be motivated one needs to believe *Necessary*: one can doubt it and still be resolute. It is rather that those who do believe *Necessary*—who believe that if they don't give up smoking now, they never will—are likely to be strongly motivated to maintain the resolution.

So far I have been concerned predominantly with a descriptive question: what is the nature of will-power and of the strength of will that it supports? Now I turn to the normative question of whether it is reasonable to have and to exercise such a faculty. In fact, although I have not been addressing it, such a question has been implicit all along. On the account presented earlier, weakness of will is, roughly, the unreasonable revision of a resolution; strength of will is its reasonable maintenance. Thus, in giving this account of will-power I have been assuming that its employment is reasonable. If it is not, then it will never result in strength of will, but simply in stubbornness.

Giving a complete account of exactly when the employment of will-power is reasonable would be a big task; I cannot embark on it here. Besides, others have already made great progress. I hope that I shall be able to incorporate much of what they say.[38] What I can do is to briefly describe how the normative issues appear when considered in the light of the account presented here, with its stress on the distinction between revision and reconsideration. In so doing I hope to sketch the lines of a defence against the obvious objection that it involves an unacceptable bootstrapping.

One natural account of the reasonableness of a procedure of practical rationality is pragmatic: people who live by it do better, in the sense of achieving their long term goals, than those who don't. We might then defend the reasonableness of will-power by pointing to the advantages gained by those who have it: advantages well documented in the empirical literature. I think that such a defence is basically right. Yet there is a worry that accompanies any such pragmatic approach. Couldn't it be the case that the world is so arranged that the unreasonable flourish? To put the point picturesquely: Couldn't there be a perverse god who rewarded the unreasonable by making sure that they flourished, and penalized the reasonable by making sure that they didn't? Then flourishing would be no indication of reasonableness.

In fact it is far from clear that such arguments work. They are most effective in showing that pragmatic advantage is no guide to *theoretical* rationality: false beliefs can be more advantageous than true. But perhaps pragmatic advantage is a good guide to *practical* rationality. Perhaps the practically reasonable thing to do in the world of the perverse god is that which brings his reward, i.e. that which would *otherwise* be unreasonable. Nevertheless, it might seem as though there is particular reason to worry about will-power, its pragmatic advantage notwithstanding.

To see the worry we need a case in which someone has formed a resolution to resist a certain temptation, and where the refusal to reconsider that resolution will be, *prima facie*, reasonable (it isn't the kind of case in which the resolution was silly and the contrary desires justifiably strong). We also want it to be a case in which, had the agent never made the resolution and were now to consider what to do, she would,

---

[38] See especially E. McClennen, *Rationality and Dynamic Choice* (Cambridge: Cambridge University Press, 1990), and for a simple presentation of his view (contrasting it with Bratman's) L. DeHelian and E. McClennen 'Planning and the Stability of Intention: A Comment', *Minds and Machines* 3 (1993) pp. 319–33. For some proposed revisions to McClennen's conception of when resoluteness is rational see D. Gauthier, 'Resolute Choice and Rational Deliberation: A Critique and a Defense', *Noûs* 31 (1997) pp. 1–25; and M. Bratman 'Toxin, Temptation and the Stability of Intention' in *Faces of Intention* pp. 58–90.

quite reasonably, choose to indulge (i.e. to do what, having formed the resolution, we describe as succumbing). In such a case it looks as though it is the resolution itself that is making all the difference: that it is the resolution itself that is making it reasonable not to succumb. But this seems to involve what Bratman has characterized as an unacceptable bootstrapping: the very fact that one has formed an intention to do something cannot by itself be a reason to do it. Otherwise we could give ourselves a reason to do something just by intending to do it; and that cannot be right.[39]

A concrete example might be helpful: a suitably qualified case of a resolution to give up smoking will serve. We imagine a regular smoker who gains pleasure and comfort from cigarettes. She is young and fit: cigarettes are not going to kill her anytime soon. Nevertheless, for the standard reasons, it is in her interest to give up smoking in the near future, and she desires to do so. So it is rational for her to form an intention to give up smoking soon. What should her attitude be towards today's habitual early morning cigarette (a cigarette which would undoubtedly set her up nicely for the day)? It seems to depend on what resolutions she has made. If she has resolved to give up next week, there is no reason to forgo today's cigarette. If she has resolved to give up today, then there is. In short: it is reasonable to give up smoking soon; having resolved to do so, it is reasonable to forgo today's cigarette, and unreasonable not to; but had she not so resolved (had she resolved instead to give up next week) it would be unreasonable to forgo today's cigarette, and reasonable not to. It seems then that it the existence of the resolution that makes all the difference. We have an instance of bootstrapping.

One response we might make is simply to insist that bootstrapping is acceptable for the special case of resolutions. Perhaps once we have resolved to do something, that does give us reason to do it. The justification comes from the need to maintain and develop the faculty of will-power, a need which does not apply to the case of intention more generally. If we fail to persist in our resolutions our faculty of will-power will be diminished. Equally importantly, since we will surely come to doubt instances of *Effective*, our confidence in the power of that faculty will be diminished.

I think that these are indeed important considerations. But they cannot give us a completely general defence of the reasonableness of will-power. For whilst it might sometimes be the case that the need to maintain and develop the faculty of will-power will outweigh the reason for succumbing to the temptation, there is no guarantee that this will generally be the case.

We need a different approach. It is provided, I think, by the distinction between reconsidering and revising a resolution. We have been talking as though the formation of an resolution provided a reason to perform the action which that resolution concerns. As such it would count in with the other reasons for and against performing that action, and hence with the reasons for and against revising the resolution. However, prior to revision comes reconsideration: if the resolution isn't reconsidered it won't be revised. And a decisive reason against reconsideration need not be a decisive reason against revision once a reconsideration is made.

---

[39] Bratman, *Intention, Plans and Practical Reason* pp. 24ff. For further discussion see J. Broome, 'Are Intentions Reasons? And How Should We Cope with Incommensurable Values?', in C. Morris and A. Ripstein (eds.), *Practical Rationality and Preference: Essays for David Gauthier* (Cambridge: Cambridge University Press, 2001) pp. 98–120.

The basic idea is this: a resolution provides one with a reason against any reconsideration that is prompted by the desires that the resolution was designed to defeat. This reason has a special status: it will only be rationally defeated in very unusual circumstances (such as the example of water abstinence, when the resolution was trivial and the desires especially, and justifiably, strong). However, once a reconsideration is under way, things change. The fact that a resolution was made before is just one consideration amongst many. Whilst it might be decisive (perhaps on the grounds of the need to maintain the faculty of will-power) it might well not be.

I suggest then that the exercise of will-power, understood as a block on reconsideration, need not be irrational in the way that it would be if we understood it as the requirement that the outcome of a reconsideration should always be the maintenance of a resolution. The rationale for will-power is indeed pragmatic; provided that it leads to pragmatic advantage in general (a condition that I have not explored), we need not fear that its bootstrapping features give rise to irrationality.

We might make a parallel with the case of a perverse god, this time one who, say, punished every correct long division calculation. One might envisage two different responses to this from the understandably aggrieved populace. The first would involve training themselves to always get the calculations wrong. That clearly would require them learning to be (at least theoretically) irrational. The second response would simply involve refraining, whenever possible, from calculating long divisions. That would put a substantial constraint on their lives, but it need involve no irrationality. Teaching yourself to always decide reconsiderations in favour of maintaining resolutions corresponds to the first of these responses; teaching yourself to avoid the reconsiderations corresponds to the second.

We might wonder whether this approach can shed light on any of the puzzle cases that have concerned writers on intention. I suspect that it can; and seeing this will shed more light on the way in which the defence of will-power is indeed pragmatic. Consider first the toxin puzzle that Gregory Kafka devised.[40] You are offered an enormous sum of money if you will form the intention to drink a toxin that will cause very unpleasant symptoms for a day, but will not otherwise harm you. Let us suppose that you judge that the benefit of the money hugely outweighs the cost of the unpleasant symptoms, and so judge it rational to form the intention to drink the toxin. However, there is a catch. You will be rewarded simply for forming the intention (as indicated by a reliable brain scanner) and your reward will come before the moment to drink the toxin arrives. Can you still rationally form the intention to drink the toxin? There is an argument that you cannot. Suppose, for *reductio*, that you could. Then, once you have received the money, it will be rational to revise your intention, since you now only stand to lose by drinking the toxin. But knowing this, it will not be possible for you to rationally form the intention in the first place.

Much debate has centered on whether or not it is rational to revise the intention once you have the money. Some have argued that, given the pragmatic advantages that forming the intention brings, it is rational to do anything that is needed in order to form it. So if one needs to avoid revising it, it is rational to avoid revising it.[41] Others

[40] G. Kavka, 'The Toxin Puzzle', *Analysis* 43 (1983), pp. 33–6.
[41] For instance D. Gauthier, 'Assure and Threaten', *Ethics* 104 (1994) pp. 690–721, at pp. 707–9.

counter that, pragmatic advantages notwithstanding, it must be rational to revise an intention whose realization will bring only costs: the best that can be said is that it is rational to make oneself irrational. [42]

On the approach suggested here, we can do justice to both of these thoughts. For there are now two questions: whether it is rational to reconsider the intention; and whether, once it is reconsidered, it is rational to revise it.[43] On the second of these questions, I side with those who argue that revision must be the rational course. The question of the rationality of reconsideration is harder. It seems that two different rules of practical rationality are engaged, and that they pull in opposite directions. You now believe that circumstances have changed in such a way as to defeat the purpose of the intention (you have the money), so you have grounds for reconsideration.[44] On the other hand, this is a resolution, and the desire to break it is the desire to avoid the unpleasant symptoms that it was designed to overcome, so in so far as there is a rational requirement to be strong-willed, there are grounds against reconsideration. We can thus understand why our reaction is so uncertain here.

We might argue that if the justification for the rules of practical reason is pragmatic, then the beneficial rule, urging non-reconsideration, should dominate. The difficulty here is that the toxin case is a one-off. It is, to say the least, unusual to meet cases of this form in daily life. So it is unlikely that there would be a general pragmatic advantage to be gained by refusing to reconsider resolutions in cases in which we believed them to be pointless. Nor, I suspect, could we rationally decide not to reconsider as a result of identifying something as a toxin case. Non-reconsideration has to be a nonreflective business, resulting from habits and tendencies that have been deeply ingrained.[45] Once we come to the point of deciding whether to reconsider it will be too late: by then we will already be reconsidering.

Nevertheless we can bring out the pragmatic rationale for non-reconsideration in cases like these by considering situations in which there would be reason and opportunity for the relevant habits and tendencies to be laid down. Suppose that we lived in an environment in which almost every decision had the form of the toxin case. Suppose that, for his own mysterious ends, the perverse god arranged things so that the necessities of life were distributed to those who intended to endure subsequent (and by then pointless) suffering. Imagine how we would bring up our children. If resolute commitment to such intentions were really the only way to form them, that is just what we would encourage. We would inculcate habits of nonreconsideration of resolutions even when they seemed pointless. Such habits would, I suggest, appear perfectly rational.[46]

---

[42] Bratman, *Intention, Plans and Practical Reason* pp. 101–6, and 'Toxin, Temptation and the Stability of Intention'.

[43] McClennen phrases his discussion in terms of the rationality of reconsideration rather than the rationality of revision: *Rationality and Dynamic Choice* pp. 227–31

[44] This was one of the rules of thumb for revision of intentions that I gave in 'Intentions and Weakness of Will'.

[45] For discussion see Bratman, *Intention, Plans and Practical Reason* p. 61

[46] Although in *Intention, Plans and Practical Reason* Bratman makes much of the distinction between the refusal to reconsider and the refusal to revise ('the two-tier theory'), he does not himself advocate this as providing a solution to the toxin puzzle. The reason he doesn't is his

A more realistic instance of this comes with another set of cases that have been much discussed, those involving reciprocity.[47] Suppose that I agree to do some onerous thing for you if you agree to do some onerous thing for me. Both of us would benefit from the exchange. Suppose that, by the nature of the case, you need to act first, and do so. I have got what I want. Why should I now bother reciprocating? But then we have a parallel worry to that which arose in the toxin case. For once you realize that I would have no reason to reciprocate, and so come to believe that I would not do so, you will not act either. So neither of us will benefit. It seems that we cannot get rational reciprocators; or, more accurately, that rational agents driven entirely by their self-interest cannot come to reciprocate in circumstances like these.

Once again I suggest that the rational agents need to develop, and get others to recognize, a tendency not to reconsider their resolutions to reciprocate. And once again I suggest that this involves no irrationality. Moreover it seems that this resonates with our moral expectations. We do not ask that those who have entered into reciprocal agreements should go on to consider whether to go through with their side of the deal, and conclude that they should. We rather ask that their compliance should come without further consideration. This is not to deny that compliance is defeasible: if something far more important comes up, or it is realized how wrong compliance would be, then it should be reconsidered. But where these factors don't intervene, reconsideration, even if it doesn't result in revision, provides an instance of a much discussed moral failing: the failing that consists in having one thought too many.

Conclusion

My contention, then, is that there is a faculty of will-power, that works, quite rationally, to block reconsideration of resolutions; and that strength of will is standardly achieved by its exercise. Moreover, this shows the Humean theory of motivation, and the theory that results from simply adding intentions to it, are false. I do not pretend to have established these conclusions with certainty. I have drawn on our subjective experience, and on empirical work in social psychology. Subjective experience is notoriously misleading; and the empirical work is sufficiently recent, and sufficiently open to alternative interpretation, that its status is not yet secure. Whether or not my conclusions stand, I hope at least to have shown that the truth in this area cannot be settled *a priori*. There is conceptual space for many competing theories; it will take a lot of work to determine which is right.

Department of Philosophy
University of Edinburgh
Edinburgh EH8 9JX   richard.holton@ed.ac.uk

---

conviction that it must be irrational to persist in drinking the toxin. See 'Toxin, Temptation and the Stability of Intention' pp. 88–9.

[47] See, for instance, Gauthier, 'Assure and Threaten'; Bratman, 'Toxin, Temptation and the Stability of Intention'; Broome 'Are Intentions Reasons?'