The Philosophical Review, Vol. 113, No. 4 (October 2004)

Rational Resolve

Richard Holton

The more thought the better, or so analytic philosophers are apt to believe. For such rationalism we are often castigated. So we might retreat to a weaker claim: at the very least, isn't it true that the more an agent thinks, the more rational they become? The contention of this paper is that even this weaker claim is too strong. There are situations in which more thought makes an agent less rational, situations in which the rational course is to think less.

The situations I have in mind involve all-too-familiar cases of anticipated temptation. We are faced with a path that we judge best and the knowledge that we risk being tempted from it. Do we respond by constraining ourselves so that we cannot yield to the temptation, or so that the cost of yielding will be greater than that of not? Doubtless we sometimes do, but these maneuvers, though the subject of much philosophical discussion, are unusual. The simplest, most direct response consists in just forming a resolution not to succumb. That is, we form an intention to stick to the best path, an intention that is explicitly designed to resist the inclinations that we predict we shall later feel. Despite their simplicity, such resolutions can be remarkably effective.

This raises two questions. One is the descriptive question of how they work. The other is the normative question of whether it is rational to persist in them. My topic here is the latter.[1] Granted that resolutions do work, can they be rational? At first sight, it seems that the answer must be yes. After all, they enable us to hold to our considered judgments against the desires that temptation engenders. Yet things are not so simple. In the first place, on desire-based accounts of rationality, it is rational to act to maximize satisfaction of one's desires, whether or not they correspond with one's judgments about what is best; and when tempted, one's desire is exactly to succumb.

So much the worse, we might think, for desire-based accounts. Surely things will be better if we move to a reason-based account on which rational agents are understood as those who act as they judge they have best reason to act.[2] Yet the problem remains. It appears that temptation typically threatens to take judgment with it, so that those who succumb not only desire to succumb, but judge that they are following the best path after all. Call this phenomenon *judgment shift*.

Those who suffer from it might be weak-willed when they abandon their resolutions, but, having revised their judgments, they are not akratic.[3]

Judgment shift is easily explained on broadly Humean accounts, where the judgment of what is best is nothing more than the projection of the strongest desire. But even accounts that hold to a more independent picture of practical judgment need to acknowledge that, as a matter of fact, it is very common. The reasons one can give to oneself for abandoning a resolution are many: the good for which one was holding out is not as good as was originally envisaged; or the reward of the temptation is greater; or succumbing just this once will do no real damage to the cause for which the resolution was formed. As Gary Watson puts it, typically when we succumb to temptation, "we are not so much over-powered by brute force as seduced" (Watson 1999, 10); and the mark of this seduction is that our judgments are affected. Empirical work in social psychology bears out this idea: when subjects yield to temptation, they tend to lower their evaluation of the good they stood to gain by holding out.[4] Of course, not every case of yielding to temptation will bring judgment shift: sometimes we judge that we are doing wrong even as we do it. But many cases will; and among these are certainly many cases in which we take resolution to be rational. So whether we take a Humean or a more cognitive approach to practical judgment (an issue that I leave open here), it will raise a problem.

Take a concrete case. Homer has not been getting much exercise, and it is starting to show. He judges, and desires, that he should do something more active. He resolves to go for a daily run, starting next Saturday morning. But as his alarm goes off early on Saturday, his thoughts start to change. He is feeling particularly comfortable in bed, and the previous week had been very draining. He could start his running next weekend. And does he really want to be an early-morning runner at all? That was a decision made in the abstract, without the realization, which now presents itself so vividly, of what such a commitment would really involve.

The case raises two challenges to the idea that it would be rational for Homer to persist in his resolution. The first is that if he were to open the question of whether it would be best to go for the run, he would undoubtedly now conclude that it would not. Succumbing to temptation would thus be in line with the judgment that he would make of what would be best. Conversely, maintaining his resolution would, it seems, be contrary to his best judgment. And, since many have

thought that acting contrary to one's best judgment must be irrational, it seems that maintaining a resolution in a case of judgment shift will be irrational. Call this the *problem of akratic resolution*. Of course, it might be contended that the judgments made under the sway of temptation are themselves irrational, and so should be discounted. Sometimes that may be right. But in many cases, Homer's included, that would be too hasty. Homer's judgments are not crazy. The bed *is* very comfortable; he *has* had a hard week. Indeed it is far from obvious that someone in Homer's situation should go for a run every morning; physical fitness is surely not a prerequisite of the good life.

This brings us to the second challenge: if it is rational for Homer to stick with his resolution, this is at least partly because he has formed it. Suppose he had decided, reasonably enough, that early morning runs were not for him: that, all things considered, he would rather go on as before and live with the consequences. It is hard to think that such a decision would be irrational. But relative to that decision, getting up early on Saturday morning to go for a run would look irrational. At the very least, there is no sense in which Homer would be rationally *required* to get up, in the way that he is so required, having made the resolution. It seems then that it is the existence of the resolution that makes all the difference. But that, in turn, seems to imply that agents can give themselves reasons for an action just as a result of resolving on that action; and that doesn't seem right. Following Bratman, call this *the bootstrapping problem*.[5]

My aim in this paper is to answer these two problems; and it is here that the idea of how it can be rational to think less comes in. To get an intuitive sense of my solution, suppose that Homer, despite his recent inactivity, is a super-resolute type. Suppose that he springs out of bed on Saturday morning, brushing aside his desire to stay in bed, and any nagging doubts about the worth of exercise, with the simple thought that he has resolved to run, and so that is what he is going to do. This changes things radically. In the first place, while it remains true that *if* he were to reconsider his resolution, he would judge it best to stay in bed, and so would be non-akratically irresolute, that is beside the point. For, since he doesn't reconsider, he doesn't form the judgment that the best thing would be to stay in bed. His judgment shift is potential rather than actual. In sticking with his resolution, he thus doesn't act contrary to his best judgment. He acts resolutely, but not akratically.

This provides the bare bones of the answer to the problem of akratic resolution: in the absence of actual reconsideration, the resolution is

not akratic after all. This solution will not extend to all cases. Sometimes agents will go on to reconsider their resolutions and will form temptation-induced judgments that they should abandon them. In such cases, sticking to the resolution will be akratic, and I shall have nothing to say to defend its rationality. But as I hope to show, to form such a judgment is to move a long way beyond simply feeling the pull of temptation. Homer, early on Saturday morning, feels a desire to stay in bed and, perhaps, has beliefs that this would cause him less harm than he once thought. However, to think this is not in itself to think that he would do better to stay in bed. Such a judgment will typically come only when he reconsiders his resolution; and it is this that he refuses to do.

Now consider the bootstrapping problem. Since Homer does not reconsider, he does not have to think that his having resolved to go for a run provides an extra reason for going for a run. Rather, it provides a reason for *not reconsidering* whether to go for a run. Insofar as he thinks there are reasons for going for a run, these are simply the reasons that led him to form the resolution in the first place. The resolution serves to entrench these reasons; it does not provide an extra one.

The key idea here is that of *rational nonreconsideration*. Homer has rational tendencies not to reconsider his resolutions, and these tendencies can confer rationality on his persistence. I am not suggesting that all resolute agents are super-resolute in the way that Homer is. But the empirical literature indicates that the approach is not far fetched. It is exactly by developing habits of nonreconsideration that agents manage to resist temptation. Moreover, even if we do not always exemplify it, the super-resolute agent provides a model that shows how sticking with a resolution can be rational, even in the face of potential judgment shift. It is rational to have a tendency not to reconsider resolutions, even in cases where, if one were to reconsider, it would be rational to change one's mind.

In proposing this account, I side with those authors who have argued for the rationality of resolute choice.[6] In the details I follow a model that has been developed by Michael Bratman for intentions more generally (Bratman 1987). Bratman calls it the *two-tier model* since it involves the assessment of the rationality of an action (the lower tier) by considering the rationality of the habit of non-reconsideration from which it follows (the higher tier); an obvious analogy is with rule utilitarianism, whereby the rightness of an act is judged by means of the rightness of the rule from which it follows. However, surprisingly to my

mind, Bratman does not endorse the extension of the two-tier model to cover the case of resolutions, that is, to those intentions that function to block temptation. On the contrary, he thinks that resolutions have a very different structure from ordinary intentions, with the result that a wholly different account of their rationality is needed (Bratman 1998). That, I aim to show, is a mistake.

Of course, it is one thing to argue that it *can* be rational to stick to one's resolutions; it is quite another to argue that it will *always* be so. The Russian nobleman forms commitments in his radical youth to philanthropic projects that he later comes to believe are worthless (Parfit 1973, 145). Is it rational for the nobleman to maintain his earlier resolution? It seems implausible that it is, however much we might find it morally praiseworthy. Or consider the pre-adolescent boy who resolves never to be susceptible to the charms of girls (Gauthier 1997). Surely, maintaining that resolution in the face of his later attraction will not be rational. We thus need some account of *when* it is rational to maintain a resolution. I suspect that nothing like a rigorous formal theory will be forthcoming. Nevertheless, the approach advocated here gives us some purchase on the circumstances in which the nonreconsideration of a resolution, and hence its maintenance, will be rational.

The Nature of Practical Rationality

We are concerned with practical rationality rather than theoretical rationality: with the rationality that governs what we do rather than what we believe. I will think of this primarily as a set of rules for action that can provide guidance for an agent, rather than as a set of standards to enable third-party evaluation.[7] It would be rather nice to start with a characterization of what practical rationality, understood in this way, is. I cannot offer that, but I will make a few remarks.

One approach characterizes the practical rationality of a rule in terms of the *outcome* that it enables one to achieve: if the outcome is beneficial, then the rule is practically rational. We can leave entirely open the nature of the benefit; we need not even assume that it must be of benefit to the agent. Then we might say that adopting the defeasible rule "stick to your resolutions" is practically rational if it enables us to achieve outcomes that are beneficial, even if we don't desire them, or judge them to be good, at the time.

Leaving aside the difficult issue about how to characterize the beneficial outcomes, it strikes me that there is something fundamentally

right about this approach. Yet there is an obvious worry that accompanies it. Couldn't it be the case that the world is so arranged that the practically irrational flourish? To put the point picturesquely: couldn't there be a perverse god who rewarded the practically irrational by making sure that they received benefits and penalized the practically rational by making sure that they didn't? Then receiving benefits would be no indication of practical rationality.

Someone might object that such arguments are effective only 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, that is, that which would *otherwise* be unreasonable.

Such a response would surely be too glib. We have an independent grip on certain principles of practical rationality, just as we have a grip on principles of theoretical rationality, and sometimes it can benefit us to violate these principles. So, for instance, people who are prepared to pursue vendettas with no regard for the cost involved might do very well in certain kinds of negotiation.[8] They are prepared to violate a certain principle of practical rationality—do not perform acts that you believe will cost you more than they benefit you—and thereby reap the benefits of a fearsome reputation. Does that make their attitude to vendettas practically rational? No; all it shows is that it can be rational to make oneself irrational.

So, discovering that following a rule is beneficial gives only prima facie grounds for saying that it is practically rational. We need to be sure that there are no principles of rationality infringed. And it is here that we confront the two problems mentioned at the outset: the problem of akratic resolution and the bootstrapping problem. They seem to show that maintaining a resolution in the face of judgment shift will typically involve one in irrationality, notwithstanding any benefit that it might bring. The time has come to consider them in a little more detail, and in particular to see if they extend to cases of mere potential judgment shift. I start with the second, the bootstrapping problem. The solution to it will bring a natural solution to the problem of akratic resolution.

The Bootstrapping Problem

Recall the worry here as it applies to intentions in general: forming an intention to do something surely cannot give one a reason to do it that one wouldn't otherwise have. If it did, we could give ourselves a reason to do something just by intending to do it; and that cannot be right. Resolutions are just a special kind of intention, so a parallel argument should apply. They too cannot give us a reason to act that we would not otherwise have. It seems then that sticking with a resolution, where one would otherwise rationally act differently, cannot be rational.

Two different responses might be made. The first looks for some special feature of resolutions, a feature that distinguishes them from ordinary intentions and that does enable them to provide extra reasons for action. I think that there is something right about this approach, but I doubt that it can provide a full answer to the bootstrapping problem. The second response, which I find far more promising, is the two-tier strategy. I take these two responses in turn.

First strategy: Resolutions furnish extra reasons

The first response holds that although bootstrapping is unacceptable for intentions in general, it is acceptable for the special case of resolutions. The idea is that once we have resolved to do something, that does give us an extra reason to do it, a reason that we can factor into any reconsideration of the resolution and that will make it rational to persist. What extra reason? One possibility is that we might simply have an overwhelming desire to persist in our resolutions, a desire that outweighs any desire to succumb to the temptation.[9] Alternatively, the reason might come from the need to maintain and develop the faculty of will-power, a need that does not apply to the case of intention more generally. We know that if we fail to persist in our resolutions, our faculty of willpower will be diminished, and that gives us a new reason to stick with any resolution that we might have made. This might be because, like a muscle, the faculty will atrophy without use. Or it might be because if we fail in our attempts to exercise it, our confidence in the faculty will decline, which in turn will reduce its effectiveness.[10]

I think that there are some important considerations here: resolutions are indeed special. But this is not enough to give us a completely general defense of the reasonableness of willpower. For a start, the picture they require if they are to provide such a defense just isn't descriptively accurate. Though most of us would doubtless prefer to be

resolute than weak, it is not true that this preference is strong enough to outweigh temptation in all cases in which persistence would be rational.[11] Nor do we always believe that by defaulting on a resolution we will massively diminish our chances of maintaining other resolutions in the future; we all know that most smokers manage to give up only after several attempts. Further, the need to preserve the faculty of willpower is only present if the faculty will be needed in the future. So, paradoxically, if I know that the rewards of one single exercise of willpower will be so great that I will not need it in the future, that will be the very time that I will be unable to exercise it. Finally, it appears that the whole approach of adding further reasons into our reconsiderations is misguided. We do not in fact manage to stick by resolutions by reconsidering them and deciding that the balance of reasons favors their maintenance. Once we get to that point it is too late. If I reconsider when the temptation has substantially skewed my judgment, it will seem to me that the resolution should be rationally revised, and thus that persistence will not display strength of will, but rather obstinacy. Obstinacy is not a faculty whose power I will want to maintain.

Could it be, however, that even if we do not go through a process of reconsideration, the factors cited here can explain why it is rational to persist? In other words: could resolutions provide extra reasons for persisting in them, even though these are not reasons that the agent will consider? This seems more plausible, but it takes us to the second, two-tier approach. For if agents do not consider the reasons, the way in which these reasons can influence their actions will be through unreflective dispositions. It is to this that I now turn.

Second strategy: the two-tier account of resolutions

The second strategy is to embrace a two-tier account, which, we have seen, is what Bratman does for the case of intentions in general. Let us follow his reasoning there. The central idea is that it can be rational to have a general policy of not reconsidering intentions in certain circumstances. This policy can confer rationality on one's action when one acts on a particular intention, rationality that that action might not otherwise have. In order to confer this rationality, Bratman convincingly argues, it must have been rational to form the intention in the first place, and it must have been rational not to revise it at each point between its formation and the time of action (Bratman 1987, 80).

Unlike the first strategy, the thought here isn't that forming an intention gives an extra reason to follow through with that intention. However, although intentions don't create new reasons for the action, they do entrench the decisions that are arrived at on the initial consideration, since they give *reasons for not reconsidering*. If the agent had not earlier considered what to do, they would now have reason to consider; but their earlier consideration provides a reason for not considering again.

The entrenchment that intentions provide is defeasible: sometimes things will change so radically from what was expected that it will be rational to reconsider the intention. However, provided things do not change radically, it will be rational to go ahead with the intention without reconsidering. This gives the possibility of what Nietzsche called the "occasional will to stupidity," since sometimes one will follow courses of action that would have seemed stupid if one were to have reconsidered (Nietzsche 1886, §107). But, by and large, not reconsidering is beneficial. It enables economy of effort (I consider once, and then do not waste scarce time and effort in further consideration); and it provides coordination advantages (having fixed an intention, my other actions and the actions of others can be coordinated around it).

It might be thought that to embrace the two-tier strategy is to accept that it is rational to make oneself irrational. That is a mistake. I would be irrational if I reconsidered an intention and decided to stick with it, even though the reasons I then had went against it. But the whole point is that there is no reconsideration; to reconsider would defeat the point of having intentions. Indeed, typically I do not even consider whether to reconsider. I simply have unreflective habits that determine when to reconsider and when not.

A more plausible line of objection is that the two-tier strategy makes our actions *arational*: since we do not reconsider, rational assessment simply does not come into it. Certainly there are ways of sticking with intentions that do involve making oneself arational. If I intend to stay in the same place for the next six hours, a powerful sleeping drug will do the job at the price of making me arational for that period. However, that is not the model that we are proposing. There are good reasons for thinking that agents who employ a strategy of nonreflective nonreconsideration do not thereby make themselves arational. First, rationality concerns what we have the *capacity* to do. In employing a habit of nonreflective nonreconsideration, we do not make ourselves unable to reconsider. We still *could* open the question up again, even if

circumstances do not change. It is just that we do not. (In developing the skill of catching a ball, I do not make myself unable to drop it.) Second, employing a habit of nonreconsideration does not involve completely closing down one's faculties. We still engage in lower-level thought about how the intention is to be implemented; and we still need to monitor to ensure that things have not changed so radically that the intention requires reconsideration after all. Although this monitoring will typically be nonreflective, it is still a rational process.

Can we apply the two-tier account to resolutions? My main contention here is that we can. The idea, of course, is that resolute agents acquire the disposition not to reconsider resolutions, even though, were they to reconsider, they would revise them. In many cases, such revisions would be rational, by the lights of the agent at the time: their judgment about what it would be best to do would have changed. Yet despite this potential judgment shift, the failure to revise would not be irrational since it would result from a policy of nonreconsideration that was itself rationally justified on pragmatic grounds. The earlier consideration, and the resolution that came from it, provide a reason for not now reconsidering.

Again it might be objected that in training oneself not to reconsider resolutions, one makes oneself arational. The issues here are exactly parallel to those for intentions in general. Certainly there are strategies for resisting temptation that involve making oneself arational; again, sleeping through the temptation is one.[12] But having the disposition not to reconsider resolutions need not be among them. It need not involve losing the capacity to reconsider; indeed, keeping oneself from reconsidering will often involve effort. Furthermore, pursuing a policy of nonreconsideration doesn't involve switching off one's mental faculties. Normal intentions, as we have seen, come with thresholds beyond which reconsideration will take place. Certainly for resolutions any such thresholds should be set very high: otherwise the corrupting effects of temptation on judgment will make the resolutions all too easily broken. Nevertheless, some such thresholds are surely needed; there is no point in persisting with one's resolution to exercise if one discovers that exercise is actually damaging one's health.[13] Equally importantly, we need to survey our resolutions to ensure that they are being implemented. This is especially so where we are trying to overcome habits—like smoking or sleeping in—that are so deeply ingrained that the actions become automatic.[14]

The Problem of Akratic Resolution

Having seen how the bootstrapping problem can be answered, we now return to the problem of akratic resolution. The problem here, recall, is that in cases of judgment shift it seems that to act resolutely will be to act akratically; and that appears irrational.

The problem of akratic resolution is an instance of a general problem about whether it can be rational to be akratic. There is little doubt that acting akratically can sometimes be the most rational course of those available: the judgments against which one acts might be crazy. The question is rather whether it nonetheless necessarily involves a degree of irrationality. Recently a number of authors have argued that it need not. To take one example: it is clear that our emotional responses can track reasons that we fail to notice in our judgments; and hence some have concluded that it can be rational to be moved by these emotions even when they run contrary to our judgments. We might, for instance, have an emotional sense that we should not trust a person, and this sense might be reliable, even though our explicit judgment is that the person is quite trustworthy.[15]

Perhaps this is right; but it is far from obvious that it is. It certainly seems as though if one makes a serious and considered judgment that a certain action is, all things considered, the best, it will involve a degree of practical irrationality to act against that.[16] It seems that this is the practical analogue of believing something when one thinks the evidence is against it; and that seems to involve irrationality, even if one's belief is true. We saw in the discussion of vendettas that it can be beneficial to be irrational. Why isn't this just another instance of the same thing? At most it seems that we have distinguished a new sense of rationality: an externalist, reliabilist sense, in which acting against one's best judgment is not irrational, to set against the internalist sense in which it is.

I cannot resolve the general issue between internalist and externalist conceptions of rationality here. What is important for us is that the two-tier account simply sidesteps the problem. For if agents do not reconsider, they do not ever form the judgment against which their resolution requires them to act. In the face of temptation they have the disposition to form those judgments, but the disposition is not realized. The judgment shift is merely potential. So they are not akratic. Moreover, this is no ad hoc solution; it is independently motivated by the need to solve the bootstrapping problem.[17]

In saying that agents do not reconsider, I do not mean that they do not think about the issue at all; as we have seen, some thought will typically be necessary for effective monitoring. Nonreconsideration requires only that they do not seriously reopen the issue of what to do, and seriously arrive at a new judgment. Nonetheless, it might seem that this makes rationality far too fragile. I am arguing that rationality can be preserved provided that the agent does not form the all-things-considered judgment that it would be best to abandon the resolution. Yet mightn't the agent form that judgment without reconsidering what to do? A little too much thought in the wrong direction, and the agent will fall over the abyss into irrationality. This in turn will mean that irrationality will be very frequent. For surely it is part of the nature of temptation that judgment shift is frequently not merely potential, but actual.

But this is to misunderstand the nature of temptation. It is certainly true that prior to any reconsideration, temptation brings new, or newly strengthened, desires. It is also true that it will bring new judgments: the judgments, for instance, that abandoning the resolution will not have some of the bad consequences previously envisaged, or that it will bring unforeseen benefits. Yet such judgments fall far short of the judgment that it would be best, all things considered, to abandon the resolution. That judgment involves not just an evaluative judgment, but a comparison: a *ranking* of one option as better than the others. And that ranking is not an abstract, impersonal one; it is a ranking of options as options for the agent. Such a ranking is not easily arrived at. It requires real mental activity from the agent. It is not the kind of thing that simply arrives unbidden.[18]

I think that this is enough to rebut the fragility worry. But I want to go further and suggest that there is an even stronger reason for thinking that we will not arrive at new all-things-considered judgments in the absence of reconsideration of what to do. How do we form all-things-considered judgments? I suggest that, standardly, we form them by deciding what to do. That is, rather than thinking that we first arrive at a judgment about what is best and then decide what to do on the basis of that judgment, things are the other way around. We start by deciding what to do and then form our judgment of what is best on the basis of that decision. This is not to say that the judgment about what is best is identical to the decision about what to do; we know that we might have made a mistake in our decision so that it does not correspond to what is best, a possibility made all the more vivid by reflecting on our own

past decisions, or those of others. It is simply that one's best way of deciding which action is best is via serious consideration about what to do.[19]

I do not claim that it is impossible to reach a judgment about what is best except via a judgment about what to do. In psychology few things are impossible. There are, for instance, reckless agents who know that their decisions about what to do are no guide to what is best; and there are depressed agents whose will is paralyzed, so that they judge what is best without being able to bring themselves to decide to do it. It is enough for my purposes if the typical, nonpathological route to best judgment is via a decision about what to do. For that will guarantee that, in the typical case, the only route to a new judgment about what is best is via a reconsideration of what to do. So if agents do not reconsider, they will not arrive at new judgments and will not be *akratic*. Rationality is even less fragile than was feared.

What of those cases in which the agent does arrive at the judgment that it would be best to succumb? This might happen, unusually, without the agent reconsidering what to do: perhaps the immediate judgment shift is so enormous that the agent can see no benefit whatsoever in persisting with the resolution. (I take it that such cases are very unusual: although temptation often leads us to believe in the advantages of succumbing, we normally retain a belief that there is *something* to be said for holding out.) Alternatively, the agent will reconsider what to do and will make a judgment that it is best to succumb, as a result of that reconsideration. In such circumstances, would persisting in the resolution involve irrationality? Addressing this takes us straight back to the general problem of the irrationality of akrasia. I suspect that it will involve irrationality: that even if persisting in the resolution is the most rational course, some local irrationality will be required if they are to get themselves out of the problem into which their revised judgment of what is best has led them.

The two–tier account thus does not ascribe rationality in every case; but it does provide a promising explanation of how maintaining a resolution will typically be rational. It is particularly attractive since it chimes so well with the empirical work on how we in fact stick by our resolutions: the primary mechanism, it seems, is exactly that of avoiding reconsideration. Once we have resolved, the best plan is to put things as far out of mind as possible. Even thinking about the benefits to be gained by remaining resolute makes an agent more likely to succumb.[20]

Bratman's Objections to the Two-Tier Account

Bratman himself declines to extend the two-tier account to the case of resolutions. He argues that the cases of ordinary intentions and of resolutions are not parallel. In some ways, this is obviously right. Typically my reason for forming a resolution is not to avoid wasting time thinking further about it, nor is it to gain coordination advantages.[21] The resolution might issue in advantages of this kind, but that is incidental. What is distinctive about resolutions, what distinguishes them from standard intentions, is that they are meant to overcome temptation. So the distinctive advantage to be gained from sticking to them is that which comes when temptation is indeed resisted. However, granting this difference does not show that the rationality of resolutions cannot be defended in the same way as the rationality of intentions. The structure is still the same: one gains a benefit by developing habits of nonreconsideration.

If the two-tier defense is not to apply to resolutions, there must be more substantial differences between resolutions and intentions. Bratman gives two. First:

> (i) We need to acknowledge that we are "temporally and causally located" agents: resolutions cannot work to overcome temptation by locking us into a strategy since we are always free to revise them; to pretend otherwise would be to engage in an irrational plan-worship (Bratman 1998, 72–73; Bratman 1999, 4).

Now it is surely true that resolutions do not work as a kind of mental binding. They cannot *force* us along a certain course of action, nor, if we are to maintain our rationality, should they be able to. However, this is a point that, I have already argued, the two-tier account can accommodate. We remain free agents, able to evaluate and revise our actions in the light of how things appear at the moment of action. Moreover, as we have also seen, this is not a way in which resolutions differ from ordinary intentions. For sticking with an intention also involves us in not reconsidering, while keeping the ability to do so. It seems then that the issue about our ability to reconsider a resolution will only be pertinent if there is reason to do so; and this brings us to Bratman's second point:

> (ii) Standardly, when we need strength of will to stick to a resolution, nothing unanticipated happens: resolutions are exactly meant to overcome *anticipated* temptation. In contrast, the standard two-tier account explains how it can be rational to maintain an intention in the face of *unanticipated* changes (Bratman 1999, 4, 8).

This second point is initially puzzling. Why doesn't the fact that there is typically no unanticipated information make it all the more reasonable to stick by one's resolution? Bratman's thought, presumably, is that in the standard cases in which it is rational to maintain an intention, one doesn't know whether one would rationally revise if one reconsidered. One would only know that if one did reconsider, and the point of the intention is to avoid such reconsideration. In contrast, in the standard cases of resolutions, one believes that if one were to reconsider at the time of the temptation, one *would* rationally revise (more precisely: the revision would be rational from the perspective of the state of mind at the time of reconsideration). This is the crux of the matter. Bratman thinks that it cannot be rational to form an intention that one believes one should later rationally revise. He endorses

> *The Linking Principle:* I shouldn't form an intention that I now believe I should, at the time of action, rationally revise.[22]

There is clearly something plausible about this principle. But it is ambiguous between

> *Weak Link:* I shouldn't form an intention that I now believe I should, at the time of action, rationally reconsider and revise;

and

> *Strong Link:* I shouldn't form an intention that I now believe that if I were, at the time of action, to reconsider, I should rationally revise.

The two-tier account of resolutions is quite compatible with Weak Link; when I form a resolution I do think that I shouldn't reconsider it in the face of temptation. The incompatibility is between the two-tier account and Strong Link. For in cases in which I expect reasonable judgment shift, I will think that were I to reconsider I would rationally revise. To get from Weak Link to Strong Link, one needs to add a principle about when it is rational to revise; something along the lines of:

> *Rational Reconsideration Principle:* If I now believe that if I were to reconsider at the time of action I would reasonably revise, then I should reconsider at that time.[23]

Once we have distinguished the two readings, we can ask where the plausibility resides. It is Weak Link that strikes me as plausible: if I think that I should reconsider and revise an intention at a later time, what

reason can I have for forming it now? In contrast, Strong Link, the principle that is incompatible with the two-tier account, is far less plausible. Indeed, I think that it is false. It is true that Strong Link isn't normally violated in standard two-tier intention cases, since in those cases, given that I don't reconsider, I don't have a belief about whether or not I would reasonably revise. But in some fairly standard intention cases, it is violated. The cases I have in mind are those in which people form an intention on the basis of imprecise information, knowing that more precise information will be available later. A traditionally militaristic example: You are defending your ship. Your instruments tell you that you are being attacked from somewhere in a 30° arc to the northeast. If you wait and calculate, you can find out the exact position of the attacker. But you are anticipating further attacks that will need your attention. Rather than waiting, finding the exact position of the attacker, and responding with a single missile, you form the intention of launching, when the optimum time comes, a barrage of missiles to cover the whole arc. In effect, you trade missiles for time to attend elsewhere.

Here it is rational not to reconsider your intention, even though your expectations about what will happen are not wrong (the attacker does come from within the arc you expect), and you believe that if you were to reconsider you would revise. Strong Link is violated. Yet this is a case of a straightforward intention that functions to economize on the time and effort that would be expended in reconsideration: exactly the kind of function that intentions should serve.

It is all the more plausible then to think that Strong Link will be violated in cases of resolutions, when the point is exactly to block reconsideration. Indeed, we can easily turn the ship-defense example into an example of a resolution by adding a few more features. Suppose that I know that I have a tendency to reopen questions that I should leave closed, thereby wasting time and decreasing my effectiveness. So I do not simply intend to fire the barrage of missiles, I *resolve* to do so, steeling myself against the temptation to reopen it that I know I will feel. When it was a simple intention, it was surely rational not to reconsider. Turning the intention into a resolution in this way cannot now make it rational to reconsider; on the contrary, if anything it makes it even more rational not to do so.[24]

I conclude then that Strong Link is false, both as applied to ordinary intentions and to resolutions; and hence that Bratman has given us no reason for not applying the two-tier approach to resolutions as well as

to ordinary intentions. I want to try to strengthen its appeal by examining Bratman's own positive account of when following through with a resolution is rational. I want to suggest that, despite his explicit rejection of the two-tier account for resolutions, his own account is best understood as a restricted version of it. This goes to show just how compelling the two-tier account is. However, once we understand Bratman's account of resolutions in this way, we will see that the restriction it imposes is not well founded. We need a much more general two-tier approach that I shall outline in the following section.

Bratman's Positive Account and the No-Regret Condition

Central to Bratman's positive account of resolutions is the *no-regret condition,* a condition on when it is rational to persist with a resolution. I meet the condition if and only if

(i) were I to stick with the resolution, then at plan's end I would be glad about it; *and*

(ii) were I to fail to stick with it, then at plan's end I would regret it.

There are two different ways of understanding the role of the no-regret condition, corresponding to the two strategies that we have examined so far. We could understand it as providing an extra reason to be factored into any reconsideration. Alternatively, we could understand it as working within a two-tier account, providing a constraint on the kinds of tendencies it would be rational to have.

Bratman's rejection of the two-tier account of resolutions suggests that he must mean the former. The condition will then work to describe rational reconsideration: as a rational agent, in reconsidering my resolutions I will decide to persist with them if they meet the condition and to abandon them if they do not. If the condition is to be factored into reconsideration in this way, then it must be one's *expectation* of regret that does the work; we cannot factor in what we do not anticipate. So the condition will have to be prefaced with a belief operator: I meet the condition if and only if *I believe* that were I to stick with the resolution I would be glad, and so on.

But why should we think that expecting that you will later regret abandoning a resolution will in general be what provides you with the additional grounds for rationally maintaining it? The problem is that if there is judgment shift, then at the moment of temptation you might

not believe that you will later regret succumbing. And even if you do, you might well not care about the later regret. You'll believe that it is unimportant, or misguided, or corrupt, and so should not influence you. Bratman acknowledges that the no-regret condition is rightly defeasible as a result of these sorts of factors: corrupt or misguided regret should not matter. What I am arguing is that if reconsideration is allowed, the belief that regret will not be felt, or that it will be misguided will mean that agents will abandon resolutions even when they should not. Of course, we could just stipulate that a person will only be rational if they have true beliefs about what and when they will regret, and if they care about avoiding it. But that is a quite unwarranted stipulation.[25]

I see no alternative but to understand Bratman's account within the context of a two-tier theory. Here it makes much better sense. The claim now is that it is rational to have a tendency not to reconsider those resolutions that meet the no-regret condition, even if, for whatever reason, this fact would not move you at the time.[26] With the condition operating in this way, we no longer need to insert the belief operator; rational tendencies are those that operate to protect you from regret, whether or not you recognize that this is what they do. If this construal is right, then Bratman's own positive account seems to entail the falsity of Strong Link and of the Rational Reconsideration Principle. The no-regret condition will often countenance maintenance of a resolution, even though I know that if I were to reconsider it I would revise it on grounds that would strike me as rational at the time.

However, once we think of it in this way, we should question how helpful the no-regret condition is. Indeed, what exactly is its role supposed to be? Bratman does not think that meeting the condition is *always* sufficient for the rational maintenance of a resolution, since we might view the regret as misplaced or corrupt. He only claims that it is *sometimes* sufficient (Bratman 1998, 87). Bratman also concedes that it is not necessary.[27] I agree. Some resolutions simply don't have a built-in end point at which the regret might be evaluated: I resolve to exercise in an ongoing way, rather than just to the end of the year. Other cases, which do have an end point, seem to call out for the maintenance of resolutions, even though they do not meet the no-regret condition.

Thus consider, and pity, Yuri. He has managed to fall in love with both Tonia and Lara. When he is with Tonia he is convinced that she

is the one and vows his undying commitment; unfortunately things are just the same when he is with Lara. Worse still, his life is so structured that he keeps spending time with each of them. As one commitment is superseded by another, and that by another, trust is lost all round. Clearly, it would be rational for Yuri to persist in his commitment to one of the women and to restructure his life accordingly; all of them recognize that. However, the no-regret condition isn't met. We can imagine him as a naturally contented type, who will not feel regret whomever he ends up with, in which case the second clause of the condition would not be met. Or we can imagine him as a naturally discontented type, who will feel regretful either way, in which case the first clause will not be met. Or we can imagine him as ambivalent, fluctuating between regret and happiness however he ends up, in which case neither clause will be stably met. Meeting the no-regret condition is not necessary for the rationality of persisting in a resolution, even for those resolutions that have an end point.[28]

If the no-regret condition is neither necessary nor sufficient, what role can we see it performing? From the perspective of the two-tier model, there is an obvious answer. The condition does not place a formal constraint on the *rationality* of persisting with an intention at all. After all, on the two-tier model it is quite possible that one will sometimes rationally perform actions that one will subsequently regret having performed: global benefit can give rise to local cost. Rather, it provides one consideration (among many) that is relevant to an assessment of the *benefit* of forming, and persisting in, a resolution. That is, its role is in diagnosing substantial rather than formal failures. Regret is a blunt tool: I can regret doing something I could never have known would be damaging; and I can regret doing what I know to be best if it still involves some harm. Nonetheless, anticipated regret is a defeasible indicator that I could do better, and as such is an *indirect* indicator of the irrationality of persistence. Let us now turn to address the question of the rationality of persistence directly.[29]

When Is Resolution Rational?

When, in general, is it rational to persist in a resolution? Since typically the decision on whether or not to reconsider will not stem from a deliberate judgment, but will follow from the operation of unconscious tendencies, this question will resolve into a question of which such tendencies are rational. So what can we say about them?

I doubt that we can say anything precise, but we can give some plausible rules of thumb that guide the different dispositions governing reconsideration of different types of resolution:

It is rational to have a tendency not to reconsider

- if one is faced with the very temptations that the resolution was designed to overcome;
- if one's judgment will be worse than it was when the resolution was formed.

It is rational to have a tendency to reconsider

- if the reasons for forming the resolution no longer obtain;
- if circumstances turn out to be importantly different from those anticipated;
- if one made an important mistake in the reasoning that led to the resolution.[30]

The obvious difficulty comes in the tension between the two sets of conditions. Cases of judgment shift will be cases where the first two rules will recommend nonreconsideration, but where the agent will believe, if he reflects on the matter, that one or more of the final three rules will recommend reconsideration. Moreover, in many cases such beliefs would be warranted. Circumstances do change; acquaintance with temptation provides new information; mistaken reasoning does come to light.

When I say that this is a difficulty, I do not mean that it is a difficulty in the account I am offering. Rather I think that the account reflects a difficulty that we have in deciding when reconsideration is in fact rational. Agents will have to learn when to put weight on the principles that favor nonreconsideration and when to put weight on those favoring reconsideration. This will be driven by knowledge of what works best; knowledge that will be different for different sorts of resolution. Resolutions concerning when to stop drinking might, for instance, need to be more insulated from reconsideration than resolutions concerning how to spend one's free time. Moreover, things will be different for different people. Those prone to self-deception will have reason to put more weight on the principles governing nonreconsideration than those who are not.

Spelling out these weightings is an exercise in practical psychology. Sometimes certain conditions will clearly not be met: the elderly Russian nobleman will, quite reasonably, be unlikely to think that he was in a better position to deliberate at the time that he made his resolution than he is at the time he comes to act on it. The same is true of the adolescent boy recalling his childish resolution to resist girls. We might say that they lack trust in their earlier selves.[31] This will lend strong support to the idea that reconsideration here is rational.[32] But in other cases it will be hard to say. Should a new study on the dangerous side effects of exercise lead me to revise my resolution to go for a daily swim? Should I postpone my resolution to give up smoking when my personal life takes an unexpected turn for the worse? Such questions will be hard to answer in the abstract, and even when all of the relevant facts are available, these questions might still resist a clear-cut answer.

Toxin Cases and Reciprocity

Much of the recent literature on intention has been concerned with the difficulties raised by Kavka's toxin puzzle (Kavka 1983). I have avoided discussion of it so far since it introduces complications that would have muddled the main lines of the account. Now, however, we are in a position to apply the account to it and to the associated issue of reciprocity. At the very least, this will provide an opportunity to show how the account is supposed to work. I hope, in addition, that the plausible treatment it affords to these cases will make it all the more convincing.

Here is Kavka's puzzle. 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 toxin's unpleasant effects, and so judge it rational to form the intention to drink it. 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 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 stand only to lose by drinking the toxin. But knowing that you will rationally revise it, it will

not be possible for you rationally to form the intention in the first place.

Let us focus 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 in order to form it one needs to avoid revising it, it is rational to avoid revising it.[33] Others counter that, pragmatic advantages notwithstanding, it must be rational to revise a resolution whose realization will bring only costs: the best that can be said is that it is rational to make oneself irrational.[34]

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.[35] On the second of these questions, I side with those who argue that revision must be the rational course. Once you reconsider, knowing that the money is in your account and that drinking the toxin will bring no further benefit, it must be rational to revise. The question of the rationality of reconsideration is harder. It seems that two of the five rules of practical rationality mentioned above are engaged and that they pull in opposite directions. You now believe that circumstances have changed in such a way that the reasons for forming the intention no longer obtain (you have the money), so you have grounds for reconsideration. On the other hand, this intention will be a resolution (a resolution not to be tempted to refrain from drinking the toxin once you have the money), and there is, as we have seen, a rational requirement to have a tendency not to reconsider resolutions in the face of the temptations that they were designed to overcome.

How might we resolve the uncertainty? We might argue that the justification for the rules of practical reason is pragmatic; that it would be beneficial not to reconsider since that would enable us to form the intention in the first place, and hence that the rule urging nonreconsideration should dominate. This would mean developing a specific tendency not to reconsider in toxin-style cases. The difficulty here is that the toxin case is a one-off. You were not brought up with similar cases; you are unlikely to meet with another. Nonreconsideration has to be a nonreflective business, resulting from habits and tendencies that have been deeply ingrained. We cannot simply decide to have a disposition not to reconsider toxin-style resolutions in order to get the money in this case. And knowing that we do not have this disposition,

it seems likely that we will not be able to form the resolution to drink the toxin at all, let alone do so rationally.[36]

Nevertheless, we can bring out the pragmatic rationale for nonreconsideration 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, a 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 their benefit had already been gained and there was only avoidable cost to come. Such habits would, I suggest, be perfectly rational, since we would go on benefiting from them.

A more realistic instance of this comes with another set of cases that have been much discussed, those involving reciprocity.[37] 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? There are, of course, moral reasons for acting. Let us suppose, however, that we are two entirely amoral creatures, moved only by considerations of our own benefit. 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 rational agents need to develop, and get others to recognize, a defeasible tendency not to reconsider their resolutions to reciprocate. And once again, I suggest that this involves no irrationality: the tendencies bring benefit to the agents concerned, but do not involve them in akratic action, or commit them to any kind of bootstrapping fallacy. Note, moreover, that this argument is not the reputation argument that is often advanced. It is not that it is rational for an agent to persist in reciprocation because it will give others reason to trust them next time round. That argument doesn't work if

there will be no next time. Rather, it is that it is rational to develop a habit of reciprocating. Although it is true that that habit loses its utility if there is to be no next time, that does not entail that we will cease to have it, nor that its employment will cease to be rational.

Summing Up

I hope that I have shown how it can be rational to stick to a resolution in the face of contrary inclinations and contrary beliefs. The mechanism involved—that of developing unreflective tendencies not to reconsider—is the same as that involved in the effective management of intentions more generally. Avoiding temptation makes use of the same mechanisms that enable us to allocate our cognitive energies wisely and to coordinate our activities over time and with others. Of course it is open to someone to say that a truly rational creature would have no use for such mechanisms; that what I have been proposing are fixes for the constitutionally irrational. Yet when we come to see what such "truly rational" creatures would have to be, we realize that they cannot provide models for us. They would not simply be immune from temptation; they would also, as Bratman has shown, be unlimited in their cognitive powers. Even then, they would lose many of the coordination benefits that we can gain. Rationality for creatures like us has to fit with the capacities and concerns that we have. It is here that rational resolve finds its place. The surprising upshot is that rationality can require us to learn when not to think.

Massachusetts Institute of Technology

References

Anthony, L. 1993. Quine as Feminist: The Radical Import of Naturalized Epistemology. In *A Mind of One's Own*, ed. L. Anthony and C. Witt, 185–225. Boulder: Westview Press.

———. 2000. Naturalized Epistemology, Morality and the Real World. *Canadian Journal of Philosophy* Supp. Vol. 26: 103–37.

Arpaly, N. 2000. On Acting Rationally Against One's Best Judgement. *Ethics* 110: 488–513.

Bratman, M. 1987. *Intention, Plans and Practical Reason*. Cambridge: Harvard University Press.

———. 1998. Toxin, Temptation and the Stability of Intention. In *Faces of Intention,* 58–90. Cambridge: Cambridge University Press, 1999.

———. 1999. Introduction to *Faces of Intention.* Cambridge: Cambridge University Press.

Broome, J. 2001. Are Intentions Reasons? And How Should We Cope with Incommensurable Values? In *Practical Rationality and Preference: Essays for David Gauthier,* ed. C. Morris and A. Ripstein, 98–120. Cambridge: Cambridge University Press.

Carver, C. and M. Scheier. 1998. *On the Self-Regulation of Behavior.* Cambridge: Cambridge University Press.

DeHelian, L. and E. McClennen. 1993. Planning and the Stability of Intention: A Comment. *Minds and Machines* 3: 319–33.

Gauthier, D. 1994. Assure and Threaten. *Ethics* 104: 690–721.

———. 1996. Commitment and Choice: An Essay on the Rationality of Plans. In *Ethics, Rationality and Economic Behaviour,* ed. F. Frain, F. Hahn, and S. Vannucci, 217–243. Oxford: Clarendon Press.

———. 1997. Resolute Choice and Rational Deliberation: A Critique and a Defense. *Noûs* 31: 1–25.

Hinchman, E. 2003. Trust and Diachronic Agency. *Noûs* 37: 25–51.

Holton, R. 1999. Intention and Weakness of Will. *Journal of Philosophy* 96: 241–62.

———. 2003. How is Strength of Will Possible? In *Weakness of Will and Practical Irrationality,* ed.S. Stroud and C. Tappolet, 39–67. Oxford: Clarendon Press.

Humberstone, I. L. 1980. You'll Regret It. *Analysis* 40: 175–76.

Jones, K. 2003. Emotion, Weakness of Will, and the Normative Conception of Agency. In *Philosophy and the Emotions,* ed. A. Hatzimoysis, 181–200. Cambridge: Cambridge University Press.

Karniol, R. and D. Miller. 1983. Why Not Wait? A Cognitive Model of Self-Imposed Delay Termination. *Journal of Personality and Social Psychology* 45: 935–42.

Kavka, G. 1983. The Toxin Puzzle. *Analysis* 43: 33–36.

Liddy, G. 1997. *Will.* New York: St. Martin's Press.

McClennen, E. 1990. *Rationality and Dynamic Choice.* Cambridge: Cambridge University Press.

McIntyre, A. 1990. Is Akratic Action Always Irrational? In *Identity, Character, and Morality,* ed. O. Flanagan and A. Rorty, 379–400. Cambridge: MIT Press.

Mischel, W. 1996. From Good Intentions to Willpower. In *The Psychology of Action,* ed. P. Gollwitzer and J. Bargh, 197–218. New York: The Guildford Press.

Moran, R. 2001. *Authority and Estrangement.* Princeton: Princeton University Press.

Nietzsche, F. 1886/1973. *Beyond Good and Evil.* Harmondsworth: Penguin.

Parfit, D. 1973. Later Selves and Moral Principles. In *Philosophy and Personal Relations,* ed. A. Montefiore, 137–69. London: Routledge and Kegan Paul.

Scanlon, T. 1998. *What We Owe to Each Other.* Cambridge: Harvard University Press.

Schelling, T. 1960. *The Strategy of Conflict.* Cambridge: Harvard University Press.

Watson, G. 1999. Disordered Appetites. In *Addiction: Entries and Exits,* ed. J. Elster, 3–28. New York: Russell Sage Foundation.

Notes

Thanks to audiences at Bristol University, Edinburgh University, MIT, Monash University, and the Research School of Social Sciences, Australian National University, to whom I presented earlier versions of this paper; and to Michael Bratman, John Broome, Rae Langton, Andrew Reisner, Jens Timmermann, and a referee for the *Philosophical Review* for comments on the written version.

[1] For some discussion of the former question, see Holton 2003.

[2] See, for instance, Scanlon 1998, chapter 1.

[3] For this distinction between weakness of will as the over-ready abandonment of resolutions and akrasia as action against best judgment, see Holton 1999.

[4] See, for instance, Karniol and Miller 1983.

[5] Bratman 1987, 24ff. For further discussion, see Broome 2001.

[6] McClennen 1990; DeHelian and McClennen 1993; Gauthier 1994, 1996, 1997. I have disagreements with Gauthier's final position that I will mention later. In contrast, I think that my position is broadly consistent with McClennen's; indeed, it might be thought of as developing philosophical underpinnings for his more formal work. One point of difference: McClennen structures his discussion in terms of the satisfaction of the agent's current and future *preferences.* I want to talk more broadly in terms of benefit, leaving it open whether this must correspond to the agent's preferences.

[7] For a discussion of the difference here, see Arpaly 2000. In saying that the rules provide guidance for agents, I do not mean that they need to formulate them explicitly, or even realize that their behavior is being regulated by them. Perhaps though, if they are really to count as agents, there must be some level on which they are endorsed. On this last point, see Jones 2003.

[8] See Schelling 1960, 16–20 for an early discussion of this.

[9] Since I'm not endorsing this possibility, I leave aside the vexed question

of whether desires can provide reasons.

[10] See Holton 2003, 55–58 for some discussion of the empirical evidence.

[11] Perhaps for a few it is. Consider the case of Gordon Liddy, who, by his own account, went in for a program of intentionally burning himself in order to build up his willpower (Liddy 1997). His resulting reputation certainly strengthened his bargaining power; though here we seem to be entering the territory in which it is rational to make oneself irrational. Thanks to Andrew Woodfield for the reference.

[12] This is the strategy used by one of the children in Mischel's delayed gratification experiments. See Mischel 1996, 202.

[13] We might here distinguish pressure for revision coming from the very inclinations that the resolutions were designed to overcome, from pressure coming from other sources, genuinely new information, for instance. Perhaps the thresholds should be sensitive only to the latter sort of pressure.

[14] For discussion of the importance of such self-monitoring, see Carver and Scheier 1998.

[15] McIntyre 1990; Anthony 1993; Anthony 2000; Arpaly 2000. For a criticism of some features of the approach of these writers (though not of the overall conclusion), see Jones 2003.

[16] For a presentation of the internal ("narrow") conception of irrationality, see Scanlon 1998, 25ff.

[17] There is an interesting question, but one that I shan't address, of how many other cases of apparent akrasia can be understood in this way.

[18] I speak of judgments, rather than beliefs, because of a strong tendency in philosophy to think of beliefs dispositionally: what one believes is what one would judge if one were to consider the matter. But that is exactly to obscure what is at issue here. These are cases in which agents would arrive at different judgments if they were to consider the matter at different times; and the question is whether they should go in for such consideration. I suspect that, in a desire to avoid a certain crude, reified picture of both beliefs and desires, philosophers have in general moved too far towards dispositional accounts. Our dispositions are simply not stable enough to support beliefs and desires understood in this way: they are far too sensitive to framing effects.

[19] There is a parallel here with the much-discussed phenomenon that one's best way of determining whether one believes that p is simply by doing one's best to determine whether or not it is the case that p. Here again, although one provides a route to the other, we recognize that the two states are different, since one's beliefs can be false. See Moran 2001, 60ff. for a nice discussion. The parallel, however, can be taken too far: in some sense, the belief case is the opposite to the case of practical deliberation. In the former, one looks to the world to discover a truth about oneself; in the latter, one looks to oneself to discover a truth about the world.

[20] Mischel 1996. For a discussion of this and of other relevant psychological results, see Holton 2003, 53–55.

[21] DeHelian and McClennen argue that sticking to resolutions can be seen as a coordination problem, once we treat the individual as a population of time slices. See DeHelian and McClennen 1993, also Gauthier 1994. Very

often the time slice asked to make the sacrifice will gain no advantage from it; these advantages will be gained only by subsequent slices.

[22] More precisely, his formulation is:

> If, on the basis of deliberation, an agent rationally settles at t_1 on an intention to A at t_2 if (given that) C, and if she expects that under C at t_2 she will have rational control of whether or not she A's, then she will *not* suppose at t_1 that if C at t_2, then at t_2 she should, rationally, abandon her intention in favor of an intention to perform an alternative to A. (Bratman 1998, 64)

Bratman puts as a constraint on rational intention formation what I am putting as an explicit injunction. For readability, I'm suppressing the reference to the availability of rational control; I assume that that is available.

[23] A rather different principle arises if we substitute "If I believe *at the time of action* that if I were to reconsider ..."; it is vulnerable to the same counterexamples.

[24] Does it make a difference that, at the time of forming the intention, although I know that I would revise it in the light of later evidence, I do not know *how* I would do so? It is true that it is this feature that makes it rational to form the intention to fire the barrage of missiles. The proponent of Strong Link might try rewriting the principle so that such cases do not fall within its scope, by requiring that the agent have a belief about how to revise:

> *Strong Link**: I shouldn't form an intention to ϕ if I now believe that if I were, at the time of action, to reconsider that intention, I should rationally intend to perform a different action ψ.

The problem with this approach is that then very many resolutions will fall outside the scope of the principle, since we will not know quite how we would respond to temptation; indeed, the resolution version of our missile example provides a case in point. The approach would thus classify some resolutions as rational and others as not on the basis of a distinction that looks utterly unimportant.

[25] There is much in common here with Bratman's own arguments against a similar suggestion in the case of ordinary intentions: if we just think of them as providing a further reason to add alongside the others, there is no guarantee that the reason is strong enough (Bratman 1987, 24).

[26] That is how I understand Bratman's own response to a similar objection raised by Tim Schroeder (Bratman 1998, 87).

[27] Bratman 1998, 98 n. 53. However, at places his argument seems to require that meeting the condition is a necessary condition for rational persistence: he holds that various cases of persistence will be irrational since they do not meet it. (His main example concerns the toxin case, to which we shall attend shortly.) Perhaps we should say, more cautiously, that he takes the no-regret condition to be the only sufficient condition yet identified, so that a failure to meet it gives prima facie grounds for a charge of irrationality.

[28] There are general reasons for thinking that the presence or absence of regret cannot be criterial for the rightness of an action. For example: I decide to bet $20 on a horse. Whatever happens, I shall regret. If the horse wins, I shall regret that I didn't stake more. If it loses, I shall regret having staked anything (Humberstone 1980).

[29] I say much the same about the proposal in Gauthier 1997 as I have said about Bratman's proposal: the conditions proposed there on when it is rational to persist with an intention are best understood on the two-tier account; but under that understanding they tell only part of the story.

[30] Conditions along these lines are proposed in Holton 1999.

[31] This idea has been interestingly explored in Hinchman 2003. Although I agree with much of what is said there, I don't take self-trust as *criterial* for rationality. Note that trusting one's earlier self exactly does not require that if one deliberated, one would come to the same beliefs, not even when the trust is explicitly factored in.

[32] Gauthier, introducing the adolescent example in his 1997, claims that it tells against a two-tier account, on the grounds that the boy's current and future desires will be better satisfied by sticking to the resolution. But the comparison that he seems to be making is between the desires satisfied by sticking with the resolution and the desires satisfied by embarking on some other strategy that renders the resolution unnecessary, such as joining a military academy that will keep him away from girls. The relevant comparison, on the account I am suggesting, is between the benefits (including desire satisfaction) to be gained by sticking with the resolution and those to be gained by reconsidering, and hence revising, it.

[33] For instance, Gauthier 1994, 707–9.

[34] See for instance the discussion in Bratman 1987, 101–6 and then in Bratman 1998. Indeed, I suspect that the conviction that drinking the toxin must be irrational, together with the thought that the two-tier account will lead to the opposite conclusion, is one of the factors that led him to abandon the two-tier account of resolutions.

[35] McClennen phrases his discussion in terms of the rationality of reconsideration rather than the rationality of revision (McClennen 1990, 227–31).

[36] The same response applies to the idea that the ideal strategy in the toxin case would be to develop an unreflective tendency that involves (i) up to the delivery of the money, thinking that one is going to persist with drinking the toxin, and (ii) once the money is delivered, reconsidering. Again, one couldn't just develop such a disposition in response to a one-off case, no matter how desirable it would be. Moreover, this idea involves further complications. First, as a matter of fact, it seems unlikely that we could ever form such a complex disposition. And even if we could, such a disposition would be bound to involve self-ignorance that would border on self-deception: one would have to believe that one was not going to reconsider when one was. In contrast, the simple habit of sticking with toxin-style resolutions could be totally transparent.

[37] See, for instance, Gauthier 1994; Bratman 1998; Broome 2001.