

WHAT IS THE LINK BETWEEN REGRET AND WEAKNESS OF WILL?

This is an Accepted Manuscript of an article published in *Philosophical Psychology* (2016), available online at <http://www.tandfonline.com/doi/full/10.1080/09515089.2015.1099621>

Mathieu Doucet, Department of Philosophy, University of Waterloo

mdoucet@uwaterloo.ca

Abstract

This paper argues a) that most contemporary accounts of weakness of will either implicitly or explicitly assume that regret is a typical or even necessary element of standard cases of weakness of will, and b) that this assumption is mistaken. I draw on empirical and philosophical work on self-assessment to show that regret need not accompany typical weak-willed behaviour, and that we should therefore revise the dominant account of the difference between weakness of will and (mere) changes of mind.

1. Introduction

Those of us who succumb to weakness of will often regret doing so. Consider an example: I resolve not to drink at the department holiday party. Once at the party, though, my colleague (who knows I work on weakness of will *and* that I love whisky, and who has a cruel sense of humor) breaks out a new bottle of 18-year-old Laphroig, and I find myself indulging despite my earlier resolution. The next morning, head pounding and mouth tasting of bog and bonfire, I reproach myself for weakly giving in to temptation and breaking my resolution.

This is an example that any philosophical model of weakness of will treats as a paradigm case, and one reason is that I later regret indulging. My action seems weak not just because I do what I had earlier intended not to do—sometimes, changing one’s mind is perfectly reasonable—but because my regret suggests that I realize that doing so was a mistake. So the idea that weakness of will normally generates regret seems, at first glance, to have strong intuitive appeal. I will argue that this is a mistake. Despite broad agreement to the contrary among otherwise competing philosophical models, weakness of will very often does *not* lead to regret. Regret involves self-assessment, and if it is to be a reliable indicator of weakness of will, such self-assessment must be *accurate*. There is abundant evidence, however, that we are often poor self-assessors, lacking in self-knowledge about important aspects of our mental lives. Moreover, the biases that block such self-knowledge are closely connected to the causes of weakness of will.

Section 2 explains the form of regret assumed to be associated with weakness of will. Section 3 shows that several otherwise competing philosophical accounts of weakness of will give regret a central role. Section 4 draws on psychological research on self-assessment to argue that this is a mistake, and that many cases of weakness of will lack regret. Section 5 argues that seemingly weak-willed regret can be present even in the absence of weakness of will. Finally, Section 6 argues that acknowledging the limits of self-knowledge, and so the ways in which weakness of will and regret can be disconnected, suggests important lessons about what precisely makes weakness of will irrational.

2. Regret

Regret is an emotion that arises in many contexts, not all of them relating to weakness of will. For example, much of the philosophical discussion of regret has focused on the rationality of

regret in moral dilemmas (e.g. Baron 1988, Marino 2001, Kahn 2011). The distinct form of regret connected to weakness of will has four characteristics. First, it is a form of “agent-regret” (Williams 1976, p. 123). If I regret drinking the whisky, then I not only judge that something bad has happened, I also judge that *I did something* I ought not to have done. Second, I judge that it would have been better for me *not* to do it. This sets weak-willed-regret apart from dilemma-regret, since in a tragic dilemma a person can regret her action without judging that it would have been better to have acted otherwise. Third, it involves the belief that my action was caused by weakness of will. This need not mean that I employ the concept ‘weakness of will’, or have a particular philosophical model in mind. It simply means that I believe that I violated my own prior commitment. I can regret doing something wrong as a result of ignorance, incompetence, or malice, and this will involve a different judgment than that involved in regretting weakness of will. Finally, this judgment is accompanied by a negative emotional reaction. Regret is not just a judgment about the causes of my action, it is also a form of self-reproach: a self-directed negative reactive attitude.

Regret is therefore much more than a vague feeling of unease or dissatisfaction. Rather, it involves a form of retrospective self-assessment. Psychological theories of regret classify it as a counter-factual emotion that relies on comparison with alternatives (e.g. Van Dijk & Zellenberg 2005). To regret one’s action is to look back on it after the fact and to pass a negative judgment on it. This means that, *if* it is true that weak-willed agents typically experience regret, two conditions must be met. First, they must typically engage in such retrospective self-assessment. Second, when they do, they must typically form *accurate* judgments about the quality and origins of their actions. I will argue that both of these are false, and so that there is no close

connection between weakness of will and regret. First, however, I show that several leading models of weakness of will are committed to the idea that there is such a connection.

3. Models of Weakness of Will

It is helpful to divide philosophical models of weakness of will into two broad categories: according to the *akratic* model, an agent is weak-willed if she acts contrary to her own best judgment about what she should do, while according to the *judgment-shift* model, she is weak-willed if she too-readily revises a certain kind of intention. Despite significant differences, both models give regret an important place in the characterization of weakness of will.

3.1 Akrasia

For Aristotle, weakness of will (*akrasia*) occurs when a person, under the influence of a strong desire for pleasure, intentionally acts contrary to her own best judgment (1999, Book VII). The weak-willed therefore judge that they ought to do one thing, and yet they do another. Aristotle's account of weakness of will has been enormously influential. It is reflected in Donald Davidson's influential account (1969) and in Al Mele's view that a "core akratic action" is one that the "agent consciously recognizes at the time of action is contrary to his better judgment" (2012, p. 8. See also Tennenbaum 2007, Ch. 7). This akratic model can be extended in various ways beyond core judgment-violation cases. For example, Mele argues that in unorthodox cases *akrasia* can also involve action contrary to an agent's *intention* rather than her judgment (2012, p. 16). Intentions and judgments are both 'practical commitments', and so Mele defends a disjunctive account: an action is weak-willed if it violates either a judgment or an intention.

Central to core cases of the akratic model is acting in a way that violates a consciously held commitment. Given this conscious awareness, some form of regret seems almost inevitable, since the weak-willed person believes that she is violating her own commitments even as she acts. Provided that her belief persists until the action is complete, she will typically satisfy the first three conditions for weak-willed regret. Her belief may not be sufficient to constitute regret, which also requires a distinct negative emotional attitude toward the belief. This emotion should, however, be a typical product of the belief that one has violated one's own commitments. An agent who felt no unease at violating what she claimed were her own standards would lead us to wonder whether they really *were* her standards. So built into core cases of akratic model is the presence of a belief that is closely linked to the four conditions for weak-willed regret.

Aristotle certainly drew a close connection, and claimed that “every incontinent is prone to regret” (1999, 1150b30).¹ It is the presence of regret that distinguishes the weak-willed agent from the vicious or intemperate one, since the latter “is bound to have no regrets” (Aristotle 1999, 1150a24). Weakness of will, on this view, is partly distinguished by a form of self-knowledge that is unavailable to the truly vicious, and it is this self-knowledge that contributes to its characteristic regret.

Some defenders of the akratic model allow for cases of *akrasia* that do not feature the violation of a conscious judgment or intention. Dylan Dodd, for example, argues that an action can be weak-willed without violating a conscious or occurrent commitment so long as it violates a policy the agent “continues to have at the time (s)he performs the action” (2007, p. 48). Such policies are revealed by our behavioural dispositions, not our conscious attitudes or explicit self-descriptions (Dodd 2007, p. 55). This means an agent can violate a standing policy without consciously recognizing that she does so. Dodd's view therefore allows for *akrasia* that lacks the

explicit self-knowledge of Mele's core cases. This might suggest that Dodd's model of *akrasia* does not build in regret. It still plays a central role, however, in determining whether a policy is in place. Whether an agent weakly acted contrary to a policy "will be shown by how he acts and feels later. Subsequent guilt feelings, promises never to do that again, and the like are all signs that the policy was still in effect and broken" (Dodd 2007, p. 54). Regret therefore plays a central evidential role in identifying the persistence of policies, and so in identifying akratic policy violations. As with 'strict' *akrasia*, Dodd's policy violation model of weakness of will therefore gives a central role to regret.

3.2 Judgment Shift

The akratic model, in its various forms, emphasizes actions contrary to existing commitments, whether conscious or not. A range of competing models of weakness of will, however, argue that weakness of will involves the over-ready *revision* of commitments, rather than the *violation* of existing commitments.

Perhaps the most prominent judgment-shift model is Richard Holton's, who identified weakness of will with the over-ready revision of a resolution.ⁱⁱ One feature of (some) intentions is that they are made in the knowledge that we will later be tempted not to carry them out. When we intend to disregard such future temptations and carry out the intention in spite of them, we form resolutions. These are intentions "designed to stand firm in the face of contrary inclinations" (Holton 2009, p. 10). When in the face of such temptations we nonetheless revise our judgments and so abandon our resolutions, we are (normally) weak-willed.ⁱⁱⁱ

A distinguishing feature of such models is that there need be no internal inconsistency, at the time of action, between the weak-willed person's commitments and her actions. This is

because, in typical cases, temptation leads to weakness of will not by *overcoming* a persisting judgment (as with the akratic model), but by *corrupting* that judgment by causing it to shift away from the original commitment (Holton 2009, p. 97). On such models, the weak-willed agent's actions and commitments are typically in alignment. It might be thought that the judgment shift model therefore breaks the connection between weakness of will and regret, since it is the agent's awareness of the conflict between action and commitment that generates regret in typical cases of *akrasia*. Lacking such a conflict, it might seem that there is no opportunity for regret. Several different versions of the judgment-shift model, however, still give regret a prominent role.

On the judgment shift model, weakness of will can occur when an agent changes her judgment about what she ought to do, but not all such changes of mind count as weakness of will. We often change our mind in the face of new information or changing circumstances. This can be perfectly rational: a refusal to change one's mind or abandon one's intentions can be a form of stubborn irrationality. So defenders of this model must provide an account of what sets weakness of will apart from rational changes of mind. There are two approaches to answering this question. The first is normative, the second descriptive. Several versions of both approaches, however, give regret a central role in accounting for weak-willed judgment shifts.

3.2.1: Normative judgment shift models

Normative approaches offer accounts of when it would be *rational* to change one's mind, and then identify weakness of will with irrational shifts in judgment. Regret can play a central role in such accounts. For example, Michael Bratman, whose account of the rationality of intentions forms the basis of Holton's own account, includes an explicit regret condition. Bratman's work

aims to make sense of how agency extends over time in a way that accounts for our capacity to plan for the future (2000). Since we can, at the same time, have both a standing self-governing policy and an occurrent desire that conflicts with that policy, one challenge is to explain which should take precedence. For Bratman, sticking to one's resolutions or policies in the face of temptation is, typically, the rational course of action, the one that the agent has most reason to perform from her own perspective. His argument that resisting temptation is rational—and so that succumbing to it is irrational—relies crucially on anticipated regret.

For Bratman, maintaining your prior intentions is rational: “(a) if you stick with your prior intention, you’ll be glad you did it; and (b) if you do not stick with your prior intention, you will wish you had” (Bratman 1998, p. 70). It is rational to stick to one's resolutions provided that both conditions are met, and (b) is a regret condition—keeping a resolution is rational if one would regret breaking it. Given the importance of cross-temporal coordination in agency, it is the anticipated regret that gives one's commitment agential authority (Bratman 2007, p. 278). This account of the rationality of resolutions indirectly gives us an account of weakness of will where regret is central. If persisting with a resolution is rational if failure to stick to it would lead to regret, then *abandoning* resolutions is *irrational* if doing so would lead to regret. It is therefore the potential presence of regret that distinguishes weakness of will from mere changes of mind. For Bratman, then, what makes a shift of judgment irrational is that it is regretted after the fact. While Bratman admits that such regret is “defeasible” (1998, p. 70), he puts it at the heart of his account of the rationality of resolution-maintenance.

Harry Frankfurt defends something similar to Bratman's no-regret condition on the rationality of intention-revision. Frankfurt distinguishes two very different sources of an inability to act on one's commitments. On the one hand, an agent may be unable as a result of psychic

forces that are “not in the fullest sense his own... and whose influence he struggles to resist” (Frankfurt 1998, p. 183). This sort of aversion sounds like *akrasia*. On the other hand, the agent may *endorse* the aversion that prevents carrying out his intention, and see it as “in the most authentic sense his own force” (Frankfurt 1998, p. 184). Frankfurt’s point in drawing this distinction is that not all cases in which someone finds himself unable to act on his judgment are irrational, since “a person’s feelings may accord better with reason than his judgment does” (Frankfurt 1998, p. 189).^{iv} What sets irrational weakness apart from rational wholeheartedness is the agent’s overall assessment of the aversion—his endorsement or rejection of the psychological force blocking him from acting on his judgment. Aversions that are endorsed are not regretted. While Frankfurt’s aim is not to explain the nature of weakness of will, his view shares the idea that the agent’s own assessment of the decision not to act on his/her commitment is central in determining whether such a decision counts as irrational and weak-willed.

Interestingly, Holton is unusual in *not* explicitly appealing to regret as a criterion of weakness of will. This is not because Holton thinks that regret is unconnected to weakness of will, however: he agrees that there is an important connection between them. Despite this connection, however, Holton rejects Bratman’s regret condition for two reasons.

First, he offers counter-examples to argue that regret cannot be either a necessary or a sufficient condition of irrational resolution revision. In some cases, a resolution can be weakly abandoned without leading to regret. Vague resolutions to do something in an ongoing way—such as a resolution to ‘exercise more in the future’—might be abandoned without the agent fully realizing and so regretting it. In other cases, the presence of regret is unavoidable, and need not indicate weakness of will. In a dilemma, *any* decision might lead to regret, but the most rational course of action might nevertheless be to commit to one horn of the dilemma.

Holton does not intend such counter-examples to suggest that regret plays no role in understanding which resolution-revisions are weak-willed, however. His point is that the regret condition cannot place a “formal constraint on the *rationality* of persisting with an intention.” If there is connection between weakness of will and regret, it is not a necessary conceptual connection. Rather, regret helps to identify potential cases of weakness of will, and so plays a role “in diagnosing substantial rather than formal failures” (Holton 2009, p. 159).

Second, Holton rejects a close conceptual link between regret and weakness of will because he defends a two-tier model of the rationality of resolution revision. Whether a particular revision was rational does not depend on the particular case. Rather, it depends on whether it reflects an overall tendency to revise in such circumstances that it would be rational to have, and a revision might lead to regret in a particular case while still reflecting rational tendency. Someone who resolves to no longer play the weekly lottery might regret her decision when her favourite numbers would have earned her \$200, while someone who abandons such a resolution and wins \$200 might not regret it at all. Given the probabilities that govern most lotteries, however, a tendency to maintain such a resolution is more rational than a tendency to revise it. As Holton puts the point, “global benefit can give rise to local cost” (Holton 2009, p. 159). Because Holton’s account of the rationality of resolution-revision has this two-tier structure, he does not rely on particular instances of regret to identify weak-willed judgment shifts.

Holton’s account of what tendencies to revise (or not) resolutions it is rational to have is intentionally vague (2009 pp. 75-6, 160-1). What he does offer is a few examples of the sorts of approximate “rules of thumb” that help determine which tendencies are rational, such as “it is rational to have a tendency not to reconsider a resolution... if one is faced with the very

temptations the resolution was designed to overcome” (Holton 2009, p. 160). Holton’s account is vague because he thinks that, in many cases, whether it would be more rational to revise or persist in a resolution is a difficult question that involves deciding what weight to give a range of competing considerations. The vagueness in the account is therefore intended to reflect vagueness in the concept the account aims to capture. As a result, he is unwilling to insist that *any* condition—including new information, or regret—is a consistent mark or rational revision or persistence. Nonetheless, he agrees that regret is both a “defeasible” and an “*indirect* indicator” of weak-willed judgment shifts (Holton 2009, p.159, emphasis in original). A plausible candidate for such a rule of thumb might therefore be ‘it is rational to have a tendency not to reconsider a resolution that one would later regret revising.’ Holton acknowledges that regret plays this role by providing “one consideration (among many) that is relevant to an assessment of the benefit of forming, and persisting in, a resolution” (2009, p. 159). So despite declining to link regret and weak-willed judgment shifts as closely as do Bratman and Frankfurt, Holton joins then in seeing regret as playing an important role in distinguishing rational from irrational shifts in judgment.

3.2.2 *Descriptive judgment-shift models*

A different, descriptive, approach is to identify weakness of will with judgment shifts brought on by a particular psychological mechanism. For example, Neil Levy argues that the shifts in judgment we call weak-willed involves changes of mind caused by a shift from System 2 to System 1 cognitive processing brought on by ego-depletion (2011). Levy’s is a deflationary argument: since the same mechanism can cause shifts we are *not* inclined to describe as weak-willed, he argues that weakness of will is not a genuine psychological kind. Rather, it is an instance of a broader psychological phenomenon. Nevertheless, his argument that the mechanism

aptly describes weak-willed judgment shifts relies essentially on the presence or absence of regret. As he puts it:

Genuinely weak judgment shifts are relatively brief and transitory... the agent typically regrets the action... It is precisely because changes of mind are produced through deliberation and are not regretted by agents that we— rightly— do not regard them as involving failures of practical rationality (Levy 2011, p. 139).

Finally, George Ainslie explains weakness of will as the result of changes of mind brought on by temporary preference inversions caused by the hyperbolic discounting of the value of future goods (2001). As he puts it, the puzzle with weakness of will is explaining how agents do something “while knowing they’ll regret it and even while trying to stop” (Ainslie 2001, p. 51). Built into the fabric of weakness of will is the idea that the judgment shift is temporary, and shifts *back*. The fact that such reversals are “avoided if foreseen... and regretted afterwards” is a “defining feature” of weakness of will (2001, p. 49).

While Levy and Ainslie’s descriptive models explain weakness of will by appeal to very different psychological mechanisms, both of them identify the target of those explanations with essential reference to a form of regret. Both descriptive models share with Bratman and Frankfurt’s normative models, with Dodd’s policy violation model of *akrasia*, and with standard strict *akrasia* models the view that retrospective self-assessment leading to regret central to weakness of will.

A wide range of otherwise very different models, then, give regret a central role in weakness of will. Interestingly, none of the accounts surveyed here— with the possible exception of Bratman’s— explicitly endorse a regret condition for weakness of will.^v What all of

the arguments for the competing accounts do, however, is appeal at crucial points to regret as an important distinguishing characteristic of weakness of will. Regret does not play precisely the same role in each account. For some, it serves as reliable evidence of the presence of weakness of will. For others, it plays a constitutive role in distinguishing weak-willed from non-weak-willed judgment shifts. But while a range of models differ dramatically on just how we ought to understand weakness of will, and on the nature of the connection between it and regret, they are united in agreeing that the connection is a close one. I argue that this is a mistake. Weakness of will commonly happens without regret, and agents can easily regret actions that are not in fact weak-willed.

4. Against Regret

If, as I argued in Section 2, the regret associated with weakness of will is characterized by an accurate retrospective self-assessment of the causes of the agent's behaviour, then the question is whether weak-willed agents do, in fact, typically engage in such clear-eyed self-assessment.

There are good reasons for doubting it, both because there are good reason reasons for doubting that *any* of us typically do so, *and* because there are additional reasons for supposing that agents susceptible to weakness of will are particularly unlikely to be accurate self-assessors. I begin by focusing the argument on the judgment shift model, since the empirical evidence on which the argument draws applies most clearly to it. The section concludes, however, by applying the argument to the akratic model as well.

There are two broad ways in which an agent can be weak-willed and yet fail to experience regret. First, an agent might engage in active, but mistaken, self-assessment; second, she might decline to engage in such self-assessment at all.

4.1 Mistaken Self-Assessment

The possibility that the weak-willed engage in mistaken self-assessment is given support by evidence from psychology that suggests that *all* of us are prone to significant failures of self-knowledge. The accurate self-assessment of genuine weak-willed regret requires agents to correctly recall both *what* they did and *why* they did it. Such recall is in fact difficult to achieve. Our memory for what we did is fallible (Neisser and Harsch 1992), and our memory for why we did it even more so: there is considerable evidence that we are quite bad at recalling both our former intentional attitudes, and the cognitive processes led us to hold our current attitudes (Nisbett & Wilson 1977, Haidt 2001, Johannson, Hall, Slikström, & Olssonet 2005, Arnold & Lindsay 2007). Much of this evidence comes from studies on the ways we explain our own behaviour, which has clear relevance in the context of weakness of will. When asked to explain why we acted as we did, it normally seems to us that we answer by engaging in introspection, and remember the reasoning that led to our actions. In fact, the evidence suggests that such cognitive processes are typically unavailable to introspection. When I explain why I did what I did, it can *seem* to me that I'm remembering, but in fact I'm often confabulating what strikes me as a plausible *post hoc* story.

This process can be accurate, since plausible stories are often true stories. The problem is that the process is less than fully reliable. In particular, it is influenced by a well-established pair of psychological biases: both the tendency of agents to interpret their previous behaviour and attitudes in a way that is *consistent* with their self-image (Swann, Rentfrow, & Guinn 2002), and bias most people display in favour of a *positive* self-image (Alicke 1985, Kunda 1990, and Kruger & Dunning 1999). We have a tendency to believe that we *always* believed, judged, or

intended what we *currently* believe, judge, or intend, and that those attitudes and actions are, *and were*, rational.

These tendencies have significant implications for the suggestion that weak-willed agents will typically experience regret. In order to recognize—and so regret—her weakness of will, an agent must recognize the inconsistency between her earlier intention and her subsequent behaviour, and must do so in a way that casts her in a negative light. The evidence, however, suggests that these tendencies can prevent us from recalling that we have changed our minds, even when we engage in active introspective self-assessment. As a result, we will fail to recognize—and so to regret—our weakness of will.

The evidence that we are prone to such error in recall is extensive, and covers a wide range of attitudes and actions. In order to show that people are prone to mistaken recall of their past attitudes, it is necessary to first measure their initial attitudes. In experimental contexts, this means that people are often asked to recall, not simply what they used to believe, but more specifically *a previously recorded answer*. That is, they are asked to recall both a specific attitude and a specific action—answering a question—rather than simply being asked to remember a vague and underspecified earlier attitude. Even when participants know that there is a record of their previous answers, their recall for previous attitudes is quite poor. In one famous study, high school students were polled about their attitudes about school bussing as a tool for racial integration, and several days later were induced (through discussion) to change their attitude. When asked to recall their earlier attitudes, the students showed a strong tendency to distort their recall of their previous attitudes to make them consistent with their current attitudes, even when reminded that there was a record of their previous attitude and encouraged to recall as accurately as possible (Goethals and Beckman 1973). In general, we have a strong tendency to

interpret our past in light of our present, and so to conclude that we *always* held the attitudes that we *currently* hold.

This effect extends far beyond political attitudes, and includes our recall of our reasons for acting (Bem and McConnell 1970), what we previously knew (Arnold and Lindsay 2007), our previous emotions (Levine 1997, Levine and Safer 2002), and even our memory of entire events (Schmolck et al 2000). Most importantly for present purposes, we are liable to misremember our *intentions* in ways that makes them consistent with our subsequent behaviour (Pieters et al. 2006). Since in cases of weakness of will our behaviour is *inconsistent* with our previous intentions, the consistency bias will tend to lead us to mistaken recall of our intentions, and so to a failure to recognize—and regret—that weakness of will.

Taken together, this evidence presents a serious challenge to the idea that weakness of will leads to regret, particularly as that idea is developed by judgment-shift models: if regret requires, at a minimum, *knowing that you changed your mind*, then in many cases, the weak-willed lack this self-knowledge. So the well-established biases in recall can serve to prevent regret from gaining a purchase, since even if we do engage in self-assessment it can prevent us from recognizing that we abandoned our resolutions and acted weakly.

One possible objection to this line of argument is that *resolutions* are highly memorable, and so relatively immune from the sorts of mistaken recall described above, perhaps because they are the result of deliberation.^{vi} But resolutions are just a kind of intention, and need not be the result of any more extensive deliberation than any other intention. In fact, Holton—whose view of weakness of will is most insistent on the importance of abandoning resolutions, as opposed to more garden variety intentions—points out that it can be difficult to determine whether an intention is a resolution, and that some resolutions are “merely implicit in an

intention” (2009 p. 77). To the extent that we can mistakenly recall our intentions, it seems likely that we can mistakenly recall our resolutions as well.

Even when we *are* aware of having changed our minds, in fact, our tendencies to biased recall can lead us to retrospectively assess the rationality of our shift in a way that preserves both our sense of internal consistency and our positive sense of self, and this too can prevent regret from taking hold.^{vii} Again, here we can see how such biases can lead the weak-willed agent to evaluate the rationality of her resolution-revision and conclude that nothing untoward occurred. This is a particular risk for weak-willed violations of vague resolutions. Someone who resolves to ‘drink less’ really does resolve to disregard contrary inclinations on many future occasions, but does not resolve to quit drinking *altogether*. This leaves room for him to rationalize any particular instance of resolution-violation as in fact consistent with the resolution to drink less. Self-deception is easier for vague concepts than for highly precise ones (Sloman et al. 2010), and since many resolutions are vague to some extent, this opens the possibility for rationalization, and so to an absence of regret in the wake of weak-willed resolutions-violations.

One way in which we can generate such rationalizations is through our tendency, when retrospectively assessing the rationality of our decisions, to engage in “post-decision consolidation through re-evaluation” (Svenson & Benthorn 1992, Svenson & Jakobson 2010). That is, we tend to confirm the rationality of our decisions by re-evaluating the options and lowering our assessment of the non-chosen option, making the decision seem even more straightforward than it was at the time. When we do so, our retrospective assessments are unlikely to be that our decision was irrational or weak-willed, and so this process will also militate against regret in cases of weakness of will.

In sum, there is ample evidence that even when we do engage in active retrospective self-assessment of weak-willed behaviour, we are liable to be misled and form mistaken assessments of our previous behaviour in ways that prevent us from feeling regret.

4.2 Avoiding Self-Assessment

The argument thus far has pointed out that active retrospective self-assessment it is not guaranteed to be accurate, and that failures of self-knowledge can block regret. But there is a second way that the weak-willed agent can fail to experience regret: she might simply fail to engage in retrospective self-assessment at all.

While we retrospectively evaluate many of our decisions, there are also many that we do not re-evaluate. In fact, several of the biases mentioned in the previous section can get a grip on our beliefs by blocking retrospective self-assessment. If I weakly abandon a resolution because I convinced myself that doing so is rational, then one way that the biases in favour of both a coherent and a positive self-image can take hold is by blocking re-assessment of my decision, and so preserving my positive and coherent self-image. If my shift in judgment was in fact weak-willed, then my failure to engage in self-assessment will prevent me from recognizing this fact, and so from experiencing regret. A tendency not to reconsider or re-evaluate decisions can contribute to a failure to recognize one's own weakness of will if it is combined with the sorts of biases or motivational irrationalities that lead to weak-willed judgment shifts.

One might think that such non-reconsideration would be relatively rare among the weak-willed. After all, judgment-shift weakness of will comes from being *too* willing to reconsider, so we might expect that the weak-willed agent would *continue* to display this over-readiness to re-assess. There are, however, reasons to think this is unlikely. If I weakly abandon a resolution,

reconsideration occurs because of the tempting motivational force of factors that I had previously resolved to ignore. Once I have succumbed to temptation, however, there need be no similar motivational force pushing me to reassess my decision. In fact, the biases in favour of consistent self-understandings and positive self-image, along with the negative implications of regret, can provide motivational force sustaining a disposition *not* to reconsider. So some weak-willed agents might be both excessive prone to reconsideration of their resolutions, and insufficiently prone to recognizing that doing so was weak or irrational once they have acted.

4.3 Against Akratic Regret

The arguments thus far have focused on the judgment shift model's close association between weakness of will and regret. It might seem, however, that the akratic model is relatively immune to such objections. After all, on the judgment-shift model regret requires an agent to engage in active retrospective self-assessment, to form a *new* judgment, and so come to believe that her action was weak-willed. As we have seen, this process can be thwarted in several ways. The akratic agent, by contrast, has not changed her judgment, so regret simply requires her to *maintain* her current belief that her action violated her commitment. Since maintaining a current belief is easier than forming a new accurate belief, regret might be common among akratics.

There are two points to make in response. First, this objection applies only to the strict version of *akrasia* in which the violated commitment is conscious at the time of action. It therefore does not apply to versions of *akrasia* that are less strict: Mele's indirect 'Socratic' *akrasia*, Dodd's policy-violation version, and any version that claims we can akratically violate a *dispositional* judgment will all be susceptible to the arguments advanced above.

Second, and more importantly, several of the psychological mechanisms that operate to block regret in judgment-shift cases of weakness of will can *also* serve to block or reduce regret in akratic cases. The akratic agent's belief that she violated her commitment will not be occurrent indefinitely: eventually, new thoughts will crowd it out. If she regrets her action for long, then that regret will depend in part on recalling her previous belief. But we have already seen that agents— even non-akratic ones— can be mistaken in the recall of both their previous mental states and in the cognitive processes that led them to action. So even a strictly akratic agent might fail to recall that she violated an occurrent commitment. Moreover, if the akratic agent does revisit her actions, both her recall of her past attitudes and her assessment of those attitudes are plausibly distorted by the twin biases in favour of a consistent and positive self-image. Regret at violating one's commitment, after all, is a negative assessment of oneself as inconsistent, and a desire to see oneself in a consistent and positive light is a motivation to revise the regret-constituting judgment.

It is true that a failure to engage in active retrospective self-assessment will not block regret in akratic cases in the way it can in judgment-shift cases. It is likely, however, that strict akratics are less likely to have a tendency not to engage in self-assessment than are those whose weakness of will arises through judgment-shift. In judgment-shift cases, the biases in favour of a positive and consistent self-image serve to preserve the agent's existing beliefs, and so militate against engaging in self-assessment. In the case of *akrasia*, on the other hand, those biases work in favour of reassessment. So while this barrier to regret in judgment shift cases does not operate in akratic cases, one of the reasons it does not is because such cases make the *other* barrier— mistaken self-assessment— more likely.

I've thus far argued that a range of cognitive biases and processes make it unlikely that weak-willed agents will consistently engage active and *accurate* retrospective self-assessment. In many ways this argument is not specific to weak-willed agents—*most* people display such biases and have trouble with accurate recall of previous mental states and processes. But the problem is not simply that the weak-willed, like the rest of us, have imperfect self-knowledge and are susceptible to mistaken self-assessment. In fact, it seems possible that the kinds of motivational biases and irrationalities that generate failures of self-assessment, and so an absence of regret, are often closely related to the very biases and irrationalities that generate weakness of will. The judgment shifts characteristic of weakness of will are irrational in part because they reveal a susceptibility to motivated rationalization, and it is the exact same susceptibility—often with very similar motivations—that generates the mistaken self-assessments that follow weakness of will.

The possibility that weak-willed judgment shifts are linked to other forms of motivated reasoning is given some support by evidence that low self-control is correlated with a wide to a range of undesirable outcomes and behaviour. For example, those with low self-control perform worse academically than those with high self-control, have lower levels of well-being, worse relationships, and lower interpersonal skills. Even more interestingly, those with high self-control are more prone to the moral emotion of guilt than to shame, while those with low self-control are more prone to shame than to guilt. (Tangney et al. 2004). Guilt tends to be associated with future directed attempts to address previous wrong-doing, such as confessing, apologizing, and making amends, while shame is associated with externalizing blame, defensiveness, and denial (Tangney 1995). Such shame responses are of the sort that can block accurate self-assessment. This evidence at least suggests that those prone to the kind of self-control failures

characterized by weak-willed judgment shifts might also display a range of other failures that block accurate self-assessments. And if weak-willed judgment shifts and mistaken self-assessments have common motivations or are generated by similar psychological mechanisms, then this gives us additional reasons to doubt that regret will typically result from weakness of will.

5. False Positives

So far, I've argued that relying on regret to identify weakness of will is likely to generate false negatives: cases where retrospective self-assessment fails and so weak-willed actions are not regretted. But perhaps the claim that regret is closely associated with weakness of will can be weakened. While we can be weak-willed without experiencing regret, it may be that when we *do* regret our actions because we judge them to have been weak-willed, such assessments are accurate. If so, this would mean that, while regret is not necessary for weakness of will, the presence of weak-willed regret is typically a very reliable indication of weakness of will. In other words, using regret as a criterion for weakness of will generates many false negatives (cases of weakness of will without regret), but rarely generates false positives (cases of regret without weakness of will).

Even this claim, however, is too strong. Retrospective self-assessment can also fail by generating false positives: cases in which a rational, non-weak-willed action is regretted as weak-willed. In other words, we can have both weakness of will without regret, and weak-willed regret without weakness of will.

For a potential example of a false positive, consider my regret at drinking the whisky. Once I arrived at the party and saw that a bottle of 18-year-old Laphroig was on offer, I

abandoned my earlier resolution not to drink. On its own, the fact that I revised by resolution does not necessarily mean that it was weak-willed, since abandoning a resolution can be rational. Perhaps the *original* resolution was folly, in which case I came to my senses in abandoning it. And even if the original resolution was rational, reconsideration can be too: for instance, if one believes that circumstances have changed in ways that defeat the resolution's purpose, then it can be rational to abandon it (Holton 1999, p. 75). Reconsideration might also be rational if one gains new relevant information that was unavailable at the time the resolution was made. The discovery that I had a rare opportunity to sample one of the world's great whiskies is potentially relevant information of this sort. If, counter-factually, I would have maintained my resolution had the only drinks on offer been domestic lager, then my decision to drink the Laphroig might be a rational response to new information. I may not have formed the resolution in the morning had I known it, because I may have thought that the benefits of indulging in the rare opportunity significantly outweighed the costs.

At this point, a defender of the connection between regret and weakness of will might point to the regret I experienced the day after drinking as evidence that my decision was weak-willed, since it seems to show that, *by my own lights*, I ought not to have abandoned my resolution. As we've seen, several judgment-shift models appeal to regret in just this way. Regret will only be *good* evidence of weakness of will, however, if the judgment emerging from my morning-after self-assessment is more rational than the judgment I formed upon reconsidering my resolution. If not, then my regret is not evidence that my decision was weak-willed.^{viii}

There are several reasons to think that the distortions and biases that are prevalent in *post hoc* assessments can generate false positives as well as false negatives. First, on either model many cases of weakness of will occur because the appeal of an imminent, smaller reward leads

the weak-willed agent to choose it over a larger, but temporally distant, reward. This process can work in reverse: the value of an already gained benefit can seem diminished when it comes time to pay for it, at which point the *cost*, and not the temporally distant benefit, is most salient.

So from the perspective of the groggy morning after, I might well regret my decision, but that assessment is not clear evidence that my decision was weak-willed, since it might reflect the mirror image of the kind of distortion of the relative value of costs and benefits that leads to judgment-shift weakness of will. This does suggest that persistent regret is *more* reliable as evidence of weakness of will than momentary regret that an agent experiences when she is paying the cost for a benefit gained earlier.

Second, there is reason to think that we will tend to regret certain actions as weak-willed whether or not they weakly violated or revised our any commitments. Folk attributions of weakness of will display two biases (REMOVED FOR ANONYMOUS REVIEW). First, people are more likely to attribute weakness of will in stereotypical contexts— such as drinking alcohol or eating— than in non-stereotypical contexts. Second, in such contexts people are more likely to attribute weakness of will with the outcome of the actions is bad than when it is good. Moreover, such attributions are common even in stereotypical or bad outcome cases that lack clear commitment violations. In other words, people have a marked tendency to attribute weakness of will in stereotypical contexts even when the action was not weak-willed, and this effect is particularly noteworthy when the outcome of the action is bad.

These findings suggest that, in stereotypically weak-willed contexts with bad outcomes, we may be liable to attribute weakness of will to *ourselves* even when our action did not involve a commitment violation. First, recall that *post hoc* explanations of our own actions typically rely on what we take to be plausible general theories for explaining behaviour, rather than accurate

retrospective access to our actual motives. Above, I argued that this can lead us to mistakenly assess weak-willed actions as rational, but for actions which we believe to be stereotypically weak-willed the process can work in reverse, thus mistakenly generating seemingly weak-willed regret without any weakness of will. Second, the quality of outcome bias might combine with the hindsight bias to generate mistaken regret: if I make my decision on the basis of the best information available to me, and nevertheless my decision turns out badly, then when I assess the rationality of my decision the hindsight bias can lead me to mistakenly judge that at the time I made the decision my judgment was irrational and so weak-willed, and the tendency to interpret bad outcomes as weak-willed might reinforce this judgment.

I have emphasized that regret involves conscious judgments about the origins and rationality of one's action. It might be objected, however, that the regret is actually a feeling of unease generated by a sub-conscious process that does not involve conscious judgments about one's action.^{ix} No doubt people have such feelings of unease after they act without engaging in active retrospective self-assessment and without forming any explicit beliefs about the origins of their actions. For such unease to play the role required by the accounts of weakness of will surveyed above, however, it would need to reliably follow weak-willed actions, and only weak-willed actions. Unease that followed actions involving difficult decisions, or actions that led to bad outcomes, would not reliably track the distinction between rational and irrational shifts in judgment. Given the existence of the outcome bias, and the tendency of even explicit assessments of the rationality of one's actions to misfire, I see no reason to suppose that such a sub-conscious tendency would be reliable in the way required by the various account of weakness of will.

6. Conclusion

Our ability to engage in accurate self-assessment is limited. As a result, we can frequently have both weakness of will without regret, and regret without weakness of will. We should therefore be hesitant about drawing a close connection between them. In making this argument, however, I do not intend to suggest that our self-understanding is hopelessly distorted, or that regret rarely or never indicates weakness of will. The claim is more modest: while the weak-willed may often experience regret, the connection between weakness of will and regret is contingent and subject to interruption. As a result, the presence of regret should not serve as a criterion for weakness of will.

Nonetheless, there will be many cases where self-assessment *is* accurate, and regret does indicate weakness of will. There may also be cases of weakness of will that are more or less likely to lead to mistaken self-assessment. In particular, weak-willed actions that we perform repeatedly might be more likely to lead to accurate self-assessment than one-off failures.^x I can more easily rationalize my drinking the whisky as consistent with a commitment to temperate living if I rarely indulge than if I do so nightly. The more often I violate my commitments, the more evidence of my weakness I will accumulate, and the more difficult such rationalization will be. Moreover, more frequent weakness will often lead to increased costs, which can in turn make regret more acute. If so, then the frequency with which I am weak-willed might well play a role in the reliability of my self-assessments.^{xi}

The existence of accurate self-assessment, however, is perfectly consistent with the argument advanced above, since my claim is simply that there are reasons to think that regret will be missing from many cases of weakness of will. As a result, we should be hesitant about drawing a close connection between them. This has particularly important implications for

judgment-shift accounts, since these rely on the presence of regret to distinguish weak-willed shifts in judgment. Given the problems with this criterion, a range of different judgment-shift accounts need to re-conceive the way in which they draw this distinction.

Normative judgment-shift models should not say that what makes a shift irrational is that it leads to regret. This leaves the assessment of the rationality of the shift at the whims of particular judgments that themselves might be the product of motivated biases. Moreover, in cases when the initial judgment shift and the failed self-assessment are linked by a common bias or distortion, it will not help to say that shifts in judgment that would be regretted in the absence of bias are weak-willed, since by hypothesis in the absence of the bias there may not have *been* a shift. After all, in cases of weakness of will we are dealing with agents who are clearly irrational to some extent—expecting that their irrationality is narrowly confined may be unrealistic. This suggests the problem with relying on the presence of regret to determine whether a judgment shift is weak-willed: it involves assessing the rationality of an agent's change of mind by deferring to the *post hoc* assessment of the agent in question, in cases where the agent has just revealed himself to be suffering from a significant level of practical irrationality. It seems unlikely that such assessments will be reliable.

The argument presented here also suggests ways in which descriptive models should be modified. They cannot argue that they have identified the right mechanism because it is one that generates regret or is correlated with regret, since the very mechanisms that generate weakness of will might very well also militate against regret or be closely associated with biases that make regret less likely. Any descriptive model, then, should justify its proposed mechanism not on the grounds that it is associated with frequent regret, but rather that it accurately captures irrationally

weak shifts in judgment, even when these do not generate regret. This means that such models will need to include a normative element.

While the development of a full model of weakness of will that does not rely on regret is beyond the scope of this paper, we can make a start at describing it in outline. Such a model must navigate a tricky middle path. On the one hand, it should preserve the sense that weakness of will involves a failure to meet the agent's very own standards and commitments. On the other hand, it should not understand this in terms of the agent's own particular in-the-moment assessment, since as we have seen such assessments are liable to a range of potential biases. This suggests that a plausible approach should involve a more systematic assessment of the agent's overall reasons for acting, even when such reasons are not consciously available to the agent at the time of self-assessment. Such an approach would bring the philosophical account of weakness of will into closer contact with the view, prevalent in social psychology, that failures of self-control—which is to say, weakness of will—involve the choice of a lower-value, short-term, locally construed and peripheral goal or interest over a higher-value, long-term, globally construed, and central goal or interest (e.g. Fujita et al. 2006). A model explaining weakness of will in terms of actions contrary to the agent's overall, long-term reasons for acting would allow for shifts in judgments that are genuinely weak-willed by the agent's very own standards, but that the agent fails to recognize as such in hindsight, and so fails to regret. Such a model would walk the required tricky middle path, and so would break the close connection between weakness of will and regret. But it would also require clearly acknowledging limits on first-person authority with respect to our own mental lives, our reasons, and our deliberative capacities.

References

- Alicke, Mark D. (1985). Global self-evaluation determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, 49, 1621-1630.
- Ainslie, George. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
- Aristotle. (1999). *Nicomachean Ethics*. 2nd edition. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999.
- Arnold, Michelle and Lindsay, Stephen. (2007) 'I remember/know/guess that I knew it all along!': Subjective experience versus objective measures of the knew-it-all-along effect. *Memory and Cognition*, 35, 1854-1868.
- Baron, Marcia. (1988). Remorse and Agent-Regret. *Midwest Studies in Philosophy*, 13, 259-281.
- Bem, Daryl and McConnell, Keith. (1970). Testing the self-perception explanation of dissonance phenomena: on the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, 14, 23-31.
- Bratman, Michael. (2007). Temptation Revisited. In *Structures of Agency*. (pp. 257-282). Oxford: Oxford University Press.
- Bratman, Michael. (2000). Reflection, Planning, and Temporally Extended Agency. *Philosophical Review*, 109, 35-61
- Bratman, Michael. (1998). Toxin, Temptation, and the Stability of Intention. In J. Coleman and C. Morris (Eds.), *Rational Commitment and Social Justice* (pp. 59-83). Cambridge: Cambridge University Press.
- Davidson, Donald. (1969). How is Weakness of Will Possible? Reprinted in E. Lepore and K. Ludwig (Eds.), (2006) *The Essential Davidson* (72-89). Oxford: Oxford University Press.
- De Ridder, D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., and Baumeister, R. Taking Stock of Self-Control: A Meta-Analysis of How Trait Self-Control Relates to a Wide Range of Behaviors. *Personality and Social Psychology Review*, 16, 76-99.
- Dodd, Dylan. (2007). Weakness of Will as Intention Violation. *European Journal of Philosophy*, 17, 45-59.
- Frankfurt, Harry. (1998) Rationality and the Unthinkable. In *The Importance of What We Care About*. Cambridge: Cambridge University Press, pp. 177-190.
- Fujita, K., Trope, Y., Liberman, N., and Levin-Sagi, M. Construal Levels and Self-Control. *Journal of Personality and Social Psychology*, 90, 351-367.

- Goethals, George and Reckman, Richard. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, 9, 491-501.
- Haidt, Jonathan. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Holton, Richard. (2009). *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
- Holton, Richard. (1999). Intention and Weakness of Will. *The Journal of Philosophy*, 95, 241-262.
- Johansson, P; Hall, L; Slikström, S; Olsson, A. (2005). Failure to Detect mismatches Between Intention and Outcome in a Simple Decision Task. *Science*, 310, 116-119.
- Kahn, Leonard. (2011). Conflict, Regret, and Modern Moral Philosophy. In T. Brooks (Ed.), *New Waves in Ethics* (pp. 7-27). New York: Palgrave MacMillan.
- Kruger, Justin and Dunning, David. (1999). Unskilled and unaware: how difficulties in recognizing one's incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134.
- Kunda, Ziva. (1999). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Levine, Linda. (1997). Reconstructing Memory for Emotions. *Journal of Experimental Psychology*, 126, 165-177.
- Levine, Linda and Safer, Martin. (2002). Sources of bias in memory for emotions. *Current Directions in Psychological Science*, 11, 169-173.
- Levy, Neil. (2011). Resisting 'Weakness of Will.' *Philosophy and Phenomenological Research*, 82, 134-155.
- Marino, Patricia. (2001). Moral Dilemmas, Collective Responsibility, and Moral Progress. *Philosophical Studies*, 104, 203-225.
- May, Josh and Holton, Richard. (2012). What in the world is weakness of will? *Philosophical Studies*, 157, 341-360.
- Mele, Alfred. (2000) *Practical Irrationality*. Oxford: Oxford University Press.
- Mele, Alfred. (2012) *Backsliding: Understanding Weakness of Will*. Oxford: Oxford University Press.
- Mele, Alfred. (1996). Socratic Akratic Actions. *Philosophical Papers*, 25, 149-159.

- Mele, Alfred. (2010). Weakness of will and akrasia. *Philosophical Studies*, 150, 391-404.
- Mischel, Walter; Shoda, Yuichi; Rodriguez, Monica. (1989). Delay of gratification in children. *Science*, 244, 933-938.
- Neisser, Ulric and Harsch, Nicole. (1992). Phantom Flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd and U. Neisser, (Eds.), *Studies of 'Flashbulb' Memories* (pp. 9-31). New York: Cambridge University Press.
- Nisbett, Richard and Wilson, Timothy. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84, p. 231.
- Schmolck, H., Buffalo E.A., and Squire L.R. (2000). Memory Distortions Develop Over Time: Recollections of the OJ Simpson Trial Verdict after 15 and 32 Months. *Psychological Science*, 11, 39-45.
- Sénémeaud, Cécile and Somat, Alain. (2009). Dissonance Arousal and Persistence in Attitude Change. *Swiss Journal of Psychology*, 68, 25-31.
- Slovic, Amos, Fischhoff, Philip, and Lichtenstein, Daniel. (2002). Self-Deception Requires Vagueness. *Cognition*, 115, 268-281.
- Svenson, Ola and Benthorp, Lars. (1992). Consolidation processes in decision making: post-decision changes in the attractiveness of alternatives. *Journal of Economic Psychology*, 13, 315-327.
- Svenson, Ola and Jakobsson, Marianne. (2010). Creating coherence in real-life decision processes: Reasons, differentiation, and consolidation. *Scandinavian Journal of Psychology*, 51, 93-102.
- Swann, William; Rentfrow, Peter; and Guinn, Jennifer. (2002). Self-Verification: The Search for Coherence. In M. Leary and J. Tangney (Eds.), *Handbook of Self and Identity* (pp. 367-384). New York: Guilford Press.
- Tangney, J. (1995). Recent Advances in the Empirical Study of Shame and Guilt. *American Behavioral Scientist*, 38, 132-145.
- Tangney, J., Baumeister, R., and Boone, A.L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, 72, 271-322.
- Tenenbaum, Sergio. (2007). *Appearances of the Good. An Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press.
- Van Dijk E., and Zeelenberg, M. (2005). On the psychology of 'if only': Regret and the Comparison between Factual and Counterfactual Outcomes. *Organizational Behavior and*

Human Decision Processes, 97, 152-160.

Williams, Bernard. (1976). Moral Luck. *Proceedings of the Aristotelian Society*, 50, 115-135.

Wilson, Anne and Ross, Michael. (2001). From Chump to Champ: People's Appraisals of Their Earlier and present Selves. *Journal of Personality and Social Psychology*, 80, 572-584.

NOTES:

Acknowledgments: Thanks to Rosalind Abdool and Eric Hochstein for valuable research assistance on this project. For helpful discussion and feedback, thanks to Susan Dimock, Kerah Gordon-Solomon, Tamar Shapiro, Abigail Scholer, Jesse Summers, and to audiences at the University of Waterloo, the University of New Brunswick Fredericton, Queen's University, Wilfrid Laurier University, The Northwestern Society for Ethical Theory, and the 2013 Canadian Philosophical Association meetings at the University of Victoria. This research was supported by the Social Sciences and Humanities Research Council of Canada.

ⁱ While Aristotle's focus is on character traits, not specific actions, those with akratic characters will a) frequently perform akratic actions, and b) typically regret those actions.

ⁱⁱ Holton has since modified his view, and now defends a view closer to Mele's disjunctive account, according to which both resolution violation and judgment violation are part of the concept weakness of will (May & Holton 2012).

ⁱⁱⁱ Mele calls such cases 'Socratic *akrasia*', and argues that they are derivatively akratic (1996). The judgment shift model differs partly in treating them as paradigmatic cases, rather than derivative ones.

^{iv} For another version of the claim that something like *akrasia* can be rational, see (Arpaly 1998).

^v Bratman's regret condition is officially offered as an account of the rationality of intention-revision, rather than of weakness of will, though the connection with weakness of will is clear.

^{vi} Thanks to an anonymous reviewer for *Philosophical Psychology* for raising the possibility of this objection.

^{vii} Not all biases are in favor of consistency: we also have a tendency to see ourselves as improving over time even when no improvement occurs (Wilson & Ross 2001). While this bias mistakenly sees the self as inconsistent when it has been consistent, the underlying motivation is the same: to promote a positive current self-image.

^{viii} The phenomenon of 'inverse *akrasia*'— actions that are contrary to an agent's judgment that are nonetheless rational— might represent a kind of false positive. The possibility of rational action against one's better judgment is defended in (Arpaly 1998). Arpaly, however, agrees that such actions really are akratic. The possibility of false positives suggests another possible explanation for such cases: the regretful agent mistakenly believes that she acts contrary to her judgment.

^{ix} Thanks to an anonymous reviewer for *Philosophical Psychology* for pressing this objection.

^x Thanks to an anonymous reviewer for *Philosophical Psychology* for suggesting this point.

^{xi} Frequent weakness might also lead to accurate self-assessment in a different way: repeated failures to stick to a resolution might lead one to conclude that one is incapable of following through, and so to abandon the resolution altogether. There may be a limit to how

frequently a person can be weak-willed before we question whether she is, in fact, violating a commitment.