

The Judgment of a Weak Will

SERGIO TENENBAUM

University of New Mexico

In trying to explain the possibility of *akrasia* (weakness of will), it seems plausible to deny that there is a conceptual connection between motivation (what one wants) and evaluation (what one judges to be good); *akrasia* occurs when the agent is (most) motivated to do something that she does not judge to be good (all things considered). However, it is hard to see how such accounts could respect our intuition that the akratic agent acts freely, or that there is a difference between *akrasia* and compulsion. It is also hard to see how such accounts could be extended to the realm of theoretical reason, but this is generally not taken to be a problem, because it is generally assumed that there is no similar phenomenon in the realm of theoretical reason. This paper argues that there is such a thing as theoretical *akrasia*, and that we can find a characterization of this phenomenon in Descartes's *Meditations*. Drawing on certain passages in the *Meditations*, we can construct an account of theoretical *akrasia*; this account can then be adapted to resolve the original problem of *akrasia* in the realm of practical reason. The account asserts that there is a conceptual connection between motivation and evaluation in free action; it also enables us to show how the akratic agent is still acting freely when he does something that he does not judge to be the best all things considered.

In the *Protagoras*, Socrates says that most people

maintain that there are many who recognize the best but are unwilling to act on it. (...) Whenever I ask what can be the reason for this, they answer that those who act in this way are overcome by pleasure or pain.¹

Both Protagoras and Socrates find this view untenable. Socrates shows that this view is untenable by first identifying pleasure with the good. He then can go on to present the following *reductio* of this popular view:

Suppose we now say that a man does evil though he recognizes it as evil. Why? Because he is overcome. By what? We can no longer say by pleasure because it has changed its name to good. Overcome we say. By what we are asked. By the good, I suppose we shall say. I fear that if our questioner is ill-mannered, he will laugh and retort: What ridiculous nonsense for a man to do evil, knowing it is evil and that he ought not to do it, because he is overcome by good.²

¹ *Protagoras*; 352e (trans. by W. K. C. Guthrie in *The Collected Dialogues of Plato*, edited by E. Hamilton and H. Cairns; Princeton: Princeton University Press, 1961).

² *Protagoras*; 355c.

Even if we find the identification of good and pleasure less palatable than Socrates' audience did, we might still accept what Kant calls the 'old formula of the schools':³

We desire only what we conceive to be good; we avoid only what we conceive to be bad.

But in this case, we seem to be bound to join Socrates' imaginary ill-mannered interlocutor in finding ridiculous the vulgar belief that we can pursue evil willingly. For, if pleasure can overcome the agent, this is because the agent desires it, and if the agent desires it then the agent has to conceive it as good. And once we make this inference it seems hard to avoid Socrates' conclusion: that people can be overcome by pleasure to pursue an evil is a ridiculous opinion. Indeed it is hard to avoid the conclusion that the opinion is as ridiculous as can be: it is a contradiction.

Yet accepting Socrates' conclusion seems to force us to deny an extremely common phenomenon. For we all have found ourselves yielding to temptation, procrastinating, eating beyond healthy limits, avoiding pain, and even abandoning the greater good for the sake of pleasure. This phenomenon is the phenomenon of weakness of the will or *akrasia*.

Since it is not easy to live with contradictions or to turn a blind eye to ordinary phenomena, it is often tempting to do away with Socrates' identification of pleasure and good or the old formula of the schools. What these two have in common is the commitment to a conceptual tie between motivation and evaluation (or a certain kind of judgment), and I will call any view committed to this kind of conceptual connection a 'scholastic view.' We can read the old formula of schools in two ways. In one way, the formula merely states a weak necessary condition for desiring something; anything that I desire I must somehow conceive to be good. On this reading, desiring (or motivation) and conceiving to be good (or evaluation) could be in important respects out of tune; for example, I could strongly desire something that I conceive to be slightly good. However, we may read the old formula of the schools as making a stronger claim; that is, that the desire for a certain thing should be *identified* with a positive evaluation of this thing. This reading rules out the above disparity between motivation and evaluation. In this paper, I will be interested primarily in the stronger reading of the formula. Indeed one of the aims of this paper will be to defend the plausibility of a scholastic view committed to the stronger reading of the formula.⁴

³ I have adapted the formulation and the denomination of the "old formula of the schools" which is found in Kant's *Critique of Practical Reason* (trans. by L. W. Beck; New York: Macmillan, 1956; p. 61). Kant himself expresses some reservations about this saying, but I will ignore them.

⁴ Of course the weaker reading will be discussed and argued for by implication. As we will see in a moment, I think the correct version of the scholastic view identifies desire with a

Socrates identifies pleasure and good because these are the things we pursue, and the formula of the school insists that we could only desire what we judge to be good. Either way, desiring, wanting or pursuing on one hand, and judging something to be good on the other hand, cannot part company. And why not find the culprit right here? If we say that the akratic agent is motivated by what he does not judge to be the best, we can accept the usual phenomenon without incurring any painful contradiction.

Yet this move creates some difficulties which we will be in a better position to assess later. But we can have a preview of the problems if we think about a different, somewhat less common, phenomenon: compulsion. I might be compelled to do something by other people, or I might be compelled to do something by my own desires, my urge getting the better of me. So a drug addict might be incapable of resisting the strength of his urge, and in a moment of madness I might indeed be *incapable* of overcoming the strength of my desire for ice cream. As Aristotle says, “when nature is the cause, no one would call the people akratic.”⁵ There seems to be a difference between *akrasia* and compulsion. Aristotle says that *akrasia* is “blameworthy and base”, and though we might find this language too strong⁶ we do conceive of an akratic person as a free and responsible agent (or, at least this is what makes the issue of *akrasia* puzzling, for there is no difficulty understanding that we might lose control of our bodily movements). Indeed, we rarely find that the claim ‘I was too weak to resist temptation’ is enough to excuse the agent from all, if any, blame. On the other hand, an agent acting under compulsion is not to blame for her actions; if my desires, working ‘behind my back’, drive me to an action despite myself I am no more free than in cases of external compulsion. One might want to deny that akratic actions are indeed free, but this might be the consequence of despairing of having failed to find room for free akratic action within one’s philosophical views. At any rate, this paper will try to explain the possibility of *akrasia* on the understanding that the akratic agent is a free agent.

Donald Davidson’s seminal article on weakness of the will tries to reconcile weakness of the will with a view of motivation in which it is identified with evaluation. In the first section of this paper, I briefly examine Davidson’s effort to resolve the problem. The second section discusses attempts to open a gap between motivation and evaluation. I will endeavor to show that the strategy of discarding the old formula of the schools is not compatible with a proper understanding of the relation between motivation and evaluation

positive evaluation *from a certain perspective*. These evaluations might be overridden upon reflection (i.e., from a reflective perspective).

⁵ *Nicomachean Ethics* (trans. by T. Irwin; Indianapolis: Hackett Publishing Co., 1985; p. 186 (1148b30)).

⁶ Aristotle himself qualifies this claim later, and saying that the akratic agent is not base but “half base” (See *Nicomachean Ethics*, 1152a15–20).

in free action. Explaining the possibility of *akrasia* in terms of a gap between evaluation and motivation seems plausible only as long as we think that this possibility is peculiar to practical reason—if we think that there is no such thing as theoretical *akrasia*. Since there is no equivalent split between evaluation and motivation in theoretical reason,⁷ if we find that *akrasia* is also possible in the theoretical realm, we have a general reason to reject this way of accounting for weakness of the will. I argue that we can find in Descartes's *Meditations* a characterization of theoretical *akrasia*. Although this is significant in itself, I argue that also we can find in the *Meditations* the basic resources to account for this phenomenon: a distinction between direct and oblique cognitions and a distinction between primary and reflective cognitions. Finally, I argue that a similar account of *akrasia* can be extended to the realm of practical reason. This account will turn out to be a scholastic account of *akrasia* that neither denies the phenomenon nor falls prey to the ills that befell Davidson's account.⁸ Although the primary aim of this paper is to explain the possibility of *akrasia*, I believe that examination of this issue will throw light on the structure of practical reason in general.

I

Davidson presents the problem of *akrasia* as the conjunction of the following principles, which are at the same time individually plausible and apparently inconsistent:

- (P1) If an agent wants to do x more than he wants to do y and he believes himself to be free to do either x or y , then he will intentionally do x if he does either x or y intentionally.
- (P2) If an agent judges that it would be better for him to do x than to do y , then he wants to do x more than he wants to do y .
- (P3) There are incontinent actions.⁹

⁷ This is not to say that there is no such thing as a motivated belief, a belief that is motivated by a certain desire. I will be ignoring these cases of motivated belief. Apart from these cases, there is little sense to be made of a parallel distinction between motivation and evaluation (more on this later). However, I will argue that, even if we ignore these cases, theoretical *akrasia* is possible.

⁸ I shall not try to explain the possibility of what David Pears calls "underivative brazen *akrasia*." (See "How Easy is *Akrasia*" *Philosophia* 11, 1982. On this issue, see also Alfred Mele's "Pears on *Akrasia* and Defeated Intentions", *Philosophia* 14, 1984) I believe that there is no difficulty in assimilating these cases to momentary changes of mind or compulsion, depending on the case, but this is a topic for a different paper. I will here assume that "underivative brazen *akrasia*" is impossible. This assumption is also implicit in Davidson's early account of *akrasia* (See "How is Weakness of the Will Possible?" in *Essays on Actions and Events*, New York: Oxford University Press, 1980).

⁹ "How is Weakness of the Will Possible?", p. 23.

(P2) clearly commits Davidson to a scholastic view. Davidson does think that we can interpret 'want' and 'judges it to be better' in such a way as to make (P2) false, but he assumes that there is at least *some* plausible interpretation of these expressions that make the combination of (P1) through (P3) paradoxical. Davidson's solution to the paradox is well known, and I will present only a brief sketch of it here, somewhat adapted to make it consistent with the terminology of the rest of the paper. Davidson argues that judgments of the form 'x is better than y' that we make when deliberating about an action are always *prima-facie* judgments relative to some consideration, the same way that judgments of probability are always relative to a certain body of evidence. Let us consider the plight of the orthodox Jewish smoker whose religion forbids him from lighting a cigarette Friday night, but feels the urge to smoke, and ends up acting on it. According to Davidson he would make the following judgments:

- (a) *Prima-facie* (from the point of view of religion), not smoking is better than smoking.

and

- (b) *Prima-facie* (from the point of view of pleasure), smoking is better than not smoking.

But of course, this is not all there is to the way that the orthodox Jew sees things. For he probably also thinks that the judgment from the religious point of view outweighs the call of pleasure, and so he judges that, all things considered, it is better not to smoke:

- (c) *Prima-facie* (from a reflective perspective), not smoking is better than smoking.

But when he acts, he needs to make an unconditional judgment. He needs to decide which is better *simpliciter*, to smoke or not to smoke. Since he acts on his desire to smoke, he makes what Davidson calls an "all-out" judgment, as follows:

- (d) Smoking is better than not smoking.

(c) and (d) characterize the orthodox Jew as an akratic agent since he acts against his judgment of what is best, all things considered. But (c) and (d) do not contradict each other, since the first is a conditional whereas the second is an unconditional judgment. (P1) and (P2) do imply that if the agent judges x to be better than y, then the agent will do x rather than y, but only if 'better than' is being used in an unconditional sense. This is compatible with the

agent finding y better than x all things considered, and thus being weak-willed.

However, this solution does not seem very satisfactory. Davidson is committed to saying that one acts from the all-out judgment, but one chooses as one's all-out judgment that which was a prima-facie judgment overridden by better claims to the good. One is struck by the mystery of how one would succumb to such an obvious mistake. If we follow Davidson's own analogy with judgments of probability, the akratic person is like the person who knows that the *New York Times* published an announcement that the presidential elections would take place today, and believes that this is a typographical error, but still goes to the local polling place. This form of stupidity deserves explanation. Davidson is aware of this problem; he presents his solution to it in the following sentences:

If r is someone's reason for holding that p , then his holding that r must be, I think, a cause of his holding that p . But, and this is what is crucial here, his holding that r may cause his holding that p without r being his reason; indeed the agent may even think that r is a reason for rejecting p .¹⁰

So, in the case of the orthodox Jew, though he did not hold that the reasons from the point-of-view of sensuous pleasure were good ones, these reasons nonetheless caused him to hold that it was better to smoke. But here Davidson seems to be paying mere lip service to the scholastic view. For in this case, the unconditional acceptance of a value judgment seems to be weirdly independent of whether the agent is ready to endorse this judgment; the judgment operates, so to speak, behind the back of the agent, despite his contrary assessment of the reasons. The fact that the desires that lead the agent to act can be said to be reasons actually has no role to play in explaining why the agent chose this particular action—as opposed to any other that can be rationalized by her beliefs and desire—since the action is not determined by the agent's overall assessment of the reasons. Davidson's characterization of akratic action can equally apply to acting on compulsion; the agent judges a certain course of action best, but some contrary desires take charge of the agent.¹¹

Since Davidson himself leans toward leaving the scholastic view behind, we might as well look at those philosophers who show no scruples in doing so. In examining these views, I hope to clarify the nature of the conceptual tie between evaluation and motivation in free action, so I can lay the ground

¹⁰ "How is Weakness of the Will Possible?", p. 41.

¹¹ For further criticisms of Davidson's views on *akrasia*, see, *inter alia*, William Charlton, *Weakness of Will* (New York: Basil Blackwell, 1988), ch. 7, and Irving Thalberg, "Questions About Motivational Strength" in Ernest LePore & Brian McLaughlin, *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (New York: Basil Blackwell, 1985).

for a plausible version of the scholastic view, and to show that preserving this tie is a *desideratum* of any account of *akrasia*.

II

Denying that there is a conceptual tie between motivation and evaluation might seem eminently plausible. We often say that something is good but not what we want, or even that we do not know what is good but we know what we want. But in another sense, it seems that the link of motivation with evaluation, and thus the scholastic formula, should be unproblematic. For, indeed, if I want something there is a sense in which I make a judgment; I take it that that which I want has some value, at least from a certain perspective. For I might regret that I wanted something, and I might judge that my want involved an illusory presentation of something as good. This might be the 'morning after' of the akratic agent, but not only of the akratic agent. I might regret having wanted something, if it turned out to be incapable of bringing me any satisfaction (the cake was stale), or if for any reason it turned out that there was no value in the object of my desire (eating a cake is after all not that satisfying, and could not justify my having wanted it so badly). In the normal cases there is no problem in identifying a want with a judgment, which might turn out to be correct or incorrect. No want or desire that I take seriously can come to me simply as a pang in my stomach, an incomprehensible yen that pushes me somewhere. Excepting some moment of insanity, my desires and wants must be directed towards certain objects, projects and aims whose point I can see. And if we see the good merely as the most abstract characterization of the aim of any practical judgment, just as we can see the true as the most abstract characterization of the aim of every theoretical judgment, the identification of motivation with evaluation seems to be unproblematic. To see some point in what I do (even if the point is just something expressed by the sentence 'I felt like doing it') and so to judge it to be good, is not a sufficient condition for free action, but it seems to be a necessary one. If I *chose* to do something, I must have *taken* it, for some reason or another, to be something worth doing. And this amounts to judging it to be good, at least in the abstract sense of 'good' described above. Of course this is not to deny that pangs or yens or even wants and desires could take control of my body and issue in a piece of behavior, but this would not be a piece of free action. If I move my hand as soon as a lit cigarette touches it, I do so in a way that does not involve any judgment; I am moved by the burn without necessarily seeing any value in this action. In this case, however, there is also no motivation involved. It would be imprecise, to say the least, to claim that the person who moves her hand in this way was moved by her desire not to be burned. These are exactly instances of our behavior that do not involve the mediating contribution of our desires.

We can think of wanting or desiring something as having a perspective from which one takes a certain course of action to be good, or worthy of being pursued. In less cumbersome language, we can say that having a desire involves seeing the point of an action. Of course, the teleological character of a desire guarantees that a desire has always a certain aim. To say that a desire is a perspective on the good, however, involves more than saying that a desire aims at something; it implies not only that a desire has an aim, but also that it presents to the agent the *point* of this aim (from a certain perspective).

We can understand better why having a desire involves seeing something as good from a certain perspective if we look at pathological cases of behavior, cases that do not seem to involve any desires. Suppose someone obsessively washes his hands every five minutes. In this case, we are ordinarily reluctant to ascribe to the agent a desire¹² to wash his hands (let alone a desire to have his hands cleaned). And the reason is that we see no point in this behavior, and assume that the agent also sees no point in it; the behavior does not issue from taking anything to be good from any perspective.

Suppose, however, one were to insist on ascribing desires to such a compulsive agent, on extending the word 'desires' also to refer to these states, states of the mind that are not connected in any way to putative evaluative judgment. Still, in the context relevant to the understanding of *akrasia*, the context of free action, we can safely assume that all desires will involve seeing something as good from a certain perspective. An agent, when acting freely, can act on a desire only insofar as this desire presents the agent with a perspective on the good; in other words, desires can contribute to free action directly, by giving us reasons to pursue their objects, only insofar as they are perspectives on the good. This point follows from a rather weak assumption about the nature of free action: that for something to count as free action it must be possible for the agent to provide some account of this action, to provide her reasons for acting in this way. An agent who had no account of her actions, whose actions were incomprehensible to herself, would not be acting freely.¹³ Now we can think of this account as a kind of deliberation that the agent could have gone through prior to the action. Whether or not we assume that the agent actually went through this deliberation, we can repre-

¹² At least, as Joseph Raz points out, not in the philosophical sense of 'desire', in which a desire is supposed to figure prominently in explanation of intentional action. Raz also defends the claim that "one can want (...) only what appears to one to be good or of value" in "The Moral Point of View" in J. B. Schneewind (ed.), *Reason, Ethics and Society* (Chicago: Open Court, 1996), p. 70ff. His argument is similar to the one I present below.

¹³ One might want to include as a free agent, an agent who acted from unconscious reasons. In this case, the agent could be incapable of accounting for her free actions. This possibility would not do any harm to my argument as long as there is *some* account to be given of her action in terms of her reasons for performing this action.

sent his reasons for acting as a hypothetical deliberation. We can now ask how a desire could enter the agent's deliberation. Could a desire that did not present its object as having any point, as being good from a certain perspective, be a putative item in deliberation? That I find myself craving something, or moved in a certain direction, is not yet a reason to do anything. Suppose I wake up in the morning and I find that I am somehow inclined to reorder the dishes, although I can't see any point in doing it: they are perfectly well-arranged and there is nothing pleasant or interesting about moving cups and plates around. When deliberating about what to do, this inexplicable craving could not give me, directly, a reason to give anything.

Of course, I might find that the craving is taking my attention away from more important tasks and decide that I better indulge this craving. Here, however, it is not the craving that is the source of these reason for the pursuit of its object, but the desire to engage in these tasks that require my craving to go away—what I want is not to reorder the dishes but to get rid of the craving. Indeed if the craving could go away by my taking a deep breath and counting to five, I'd probably prefer this alternative, since the craving gives me no reason to reorder the dishes. Quite the contrary, if I would continue to find myself thus inclined, rather than thinking that reordering my dishes should be given a certain weight in my deliberations, I would look for professional help. To the extent that this urge takes hold of me, I am no longer acting freely; in fact, I am exhibiting a paradigmatic case of compulsive behavior. As Joseph Raz puts it:

So if I want to count the blades of grass in my garden I do so because I think that this will take my mind off some upsetting event, or because the action has some other good-making property. I find myself drawn to count blades of grass, but cannot think of any reason for doing so, I would certainly deny that this is a desire of mine. It is a force which seizes me in spite of myself. If I am overcome by it and perform the action, I would be right to say that I could not help it.¹⁴

However, the case for a scholastic view is not so simple. There are good reasons not merely to identify the things we are motivated to do with the things we value or judge to be good. I propose to proceed as follows: I will examine Gary Watson's contention that motivation and evaluation may be, in certain cases, *completely* divorced.¹⁵ This will serve a twofold purpose. On the one hand, it will make it clear why the conceptual tie between motivation and evaluation cannot be simply one of mutual implication. We must concede to Watson that we do not judge to be good all the things that we desire.

¹⁴ "The Moral Point of View", p. 71.

¹⁵ "Free Action" *Journal of Philosophy* v. 57 (1975), pp. 205–20. Note that Watson himself does not try to make use of this point in order to account for how it is possible to act freely against one's best judgment. See his "Skepticism about Weakness of the Will" *Philosophical Review* 86 (1977).

On the other hand, this discussion will also show why a non-scholastic analysis of these cases, such as Watson's, is inadequate.

Watson argues that there is no sense in which we value the objects of certain desires we have, or judge the objects of these desires to be good.¹⁶ Watson gives us the example of a squash player who has a sudden urge to smash a racket on her opponent's head in frustration. It would indeed be highly misleading to say that this person sees some value in smashing the racket on her opponent's head, but that this value is overridden by other values. The agent sees *no* value in this action, not even a smidgen of value. Although this point is undeniable, it should not lead us to go too far toward severing the ties between motivation and evaluation. This example only shows that such an agent, *from a reflective perspective*, does not attribute any value to such a desire, indeed, not even a smidgen of value. This is not an objection, however, to the claim that desires are always claims to or perspectives on the good, and thus judgments of the good *from a certain perspective*. In the above case, the reflective judgment takes the claim to be merely illusory. Similarly, in the realm of theoretical reason we can distinguish between a consideration that *appears* to make a claim plausible and a consideration that *in fact* lends some plausibility to a claim. A consideration pertaining to the former category will not necessarily pertain to the latter. For instance, one often finds appealing the form of reasoning which urges that one is less likely to win the state lottery if one simply chooses the same numbers that were drawn in the previous week, given that it is very unlikely that the same number will be the winning number twice in a row. But the gambler's fallacy does not yield any plausibility to a belief, not even a smidgen of plausibility, despite the fact that it is quite hard to free oneself from the illusion of validity. Suppose we want to spell out the conceptual connection between our being inclined to believe a certain judgment, and a notion such as validity or truth. The above example shows that one might be inclined to believe¹⁷ a form of reasoning that one (reflectively) judges to be wholly invalid. But one should not therefore conclude that there is no conceptual connection in this area; we merely placed it at the wrong location. For it is only insofar as I take the gambler's fallacy to *appear* to be valid that I am inclined to believe it. The same goes for the conceptual connection between

¹⁶ In a later paper, "Free Action and Free Will" (*Mind* 96, 1987), Watson distinguishes between valuing a certain thing and judging it to be good. The distinction does not appear in "Free Agency" and I doubt its intelligibility, so I will just use the two expressions interchangeably.

¹⁷ Belief and desire are often taken to be analogous terms. But this is not true, since 'belief' is an 'all-out' term (I cannot, insofar as I am rational, believe contradictory statements), whereas 'desire' isn't (I can desire incompatible states-of-affairs, without failure of rationality). The closest practical counterpart of belief is intention, and the closest theoretical counterpart to desire is an inclination to believe something.

desiring and judging to be good. It is only insofar as something appears to be good that one desires it.

This point is well illustrated by another of Watson's examples. Watson compares the man who gives up sex because he "decides that the most fulfilling life for him will be one of abstinence", and the man who "thinks his sexual inclinations are the work of the devil".¹⁸ Watson thinks that to understand this difference we must take this second man to be motivated by his sexual inclinations while not seeing any value in them. This might be true, but this does not mean that we cannot understand his desire as a perspective on the good, even if he deems this perspective illusory. This is to say that agent *has* the perspective (his sensuous nature provides him with this perspective on the good), but does not endorse it. Compare this with a person who sees the two lines in the setup of the Müller-Lyer illusion, but knows they are of the same size. This person is still, in a sense, under a perceptual illusion. The drawings still *appear* to her as of different sizes, but she knows better than to endorse the perspective afforded by this illusion.

The moderate celibate, on the other hand, does not consider the perspective of his sexual desire illusory. Rather, he takes it to provide him with insight into a form of value which is overridden by other values. Watson's description, in fact, fails to capture the predicament of the devil-fearing celibate. For Watson thinks that an action done against our evaluation is not free. But the true work of the devil is to induce a human being into sin of her own accord. Because compulsion isn't sin, Watson's devil does no better than the likes of Stalin. The real devil lets the celibate be the agent of his own evil; he makes the celibate *take* something evil *to be good*.

I do not wish to deny that motivation and evaluation might just come completely apart. But when this happens we seem to face a case of compulsion. If we take what happens to an addict, or a kleptomaniac, whose desire to have one more shot of vodka or to steal is one that he wishes he could act against, we do have a case in which the agent's judgments and evaluations are completely powerless. But these cases are different from the case of *akrasia*, exactly because of the way in which the motivation of the agent bypasses her judgment, making her unfree. It is important not to misdescribe this case as a case in which the agent's desire did not involve a judgment, as if it did not involve perceiving the action she performs as good (or desirable) from a certain perspective. When motivation and evaluation come apart, the action is not consequent, causally or inferentially, upon this judgment. Rather, the action bypasses this judgment, and the agent does not act freely or responsibly. But if our views on *akrasia* are motivated by the desire to preserve an ordinary understanding of this phenomenon, we need to explain why we hold the akratic agent to be just as (or almost as) responsible as the agent who acts

¹⁸ *Free Agency*, p. 210.

according to her better judgment, and why we do not make excuses for him in the same way that we make excuses for agents who act under compulsion.

Alfred Mele, for instance, tries to explain the phenomenon of *akrasia* as a case in which the strength of motivations does not correspond to the strength of evaluations; that which the agents judges best is not what the agent is most motivated to do.¹⁹ But he says that this explanation does not reduce *akrasia* to compulsion since it does not rule out the possibility that the akratic agent is still capable of exercising self-control. The agent might engage in a certain action that will make the strength of motivations equal the strength of evaluation. However, this seems to push the problem one step further. We now need to ask why the akratic agent failed to engage in self-control, and the same questions about weakness of the will should resurface. In order to see that this appeal to self-control does not work, it is worth examining Mele's view in more detail, especially his attempt to answer a similar criticism raised by Watson.²⁰

According to Watson, if an agent chooses freely not to exercise self-control, the agent must have changed his judgment. That is, if the agent could have exercised self-control in such a way that he would choose A over B, and yet decides not to do it, then it could no longer be true that the agent still holds A to be better than B. It is exactly this point that Mele contests:

Perhaps to choose not to implement a *choice* would be to abandon the choice; but it does not follow that one who chooses not to exercise self-control in support of one's better judgment no longer holds that *judgment*. We may, without obvious contradiction, describe a case in which an agent judges that all things considered it is better to do A than B, but due in part to his taking his reasons for doing A to be only slightly more weighty than his reason for doing B, decides to indulge himself and to refrain from exercising self-control in support of A. In such a case, the agent may think his doing B to be *permissible*, even though he judges A to be better; and he may self-indulgently opt for the lesser alternative.²¹

According to Mele, Watson has failed to show that the person who chooses not to exercise self-control in a way that makes it possible for her to choose A over B has thereby given up her judgment that A is better than B. Mele describes a case in which an agent thinks that his moral qualms about strip-tease clubs provide him with a "slightly better reason, all things considered, not to enter the clubs than to enter them",²² but he also experiences a certain desire to enter one of the clubs. He has a "desire-eradicating device" that he could use to do away with this desire, thus making sure that he would

¹⁹ See his *Irrationality* (New York: Oxford University Press, 1987). See also John Heil's "Minds Divided" (*Mind* 98, 1989).

²⁰ For Watson's version of the criticism, see "Skepticism about Weakness of the Will". Mele's response can be found in "Is Akratic Action Unfree?" (*Philosophy and Phenomenological Research*, XLVI, 1986), and *Irrationality*, ch. 2.2.

²¹ *Irrationality*, p. 28.

²² *Irrationality*, p. 28.

stay away from the clubs. According to Mele, however, there is nothing self-contradictory about imagining that the agent will choose not to use his desire-eradicating device, while still judging that, all things considered, it is better not to enter the club. It is clear why Mele chooses a case in which the agent thinks there is only a 'slightly better reason' to take a certain course of action, and that he stipulates that the agent will think that "his doing B is permissible". It is much harder to make the case if we assume that the agent thought that A was overwhelmingly better than B, or that B was not permissible.²³ I will first examine why Mele will face problems if he tries to provide an example of this kind. This will help us understand why Mele's response to Watson is unsatisfactory even in the case he describes.

Let us assume that Joe lives in a dictatorship, and the dictators are after his sister who is the head of an underground guerrilla group, and Joe knows that they will show no mercy towards his sister. The government offers Joe a million dollars to turn her in and he refuses, because he finds it abominable to trade his sister's life for a few bourgeois comforts. The government explains to Joe that the offer will still be standing for 48 hours, and that they will bring to his home tonight a suitcase full of cash, hoping he'll have changed his mind by then. They'll open it in front of him, grin and wait; and Joe knows he cannot resist such a vivid display of hard currency. Fortunately, Joe can avoid facing this irresistible temptation by just putting a sign on the door that says: "Joe does not live here anymore". Not being very bright, the government agents will turn around and never come back.

Could Joe just choose not to put a sign on the door, and yet retain his judgment that he should not turn in his sister? Now it seems that the answer is 'no', and that is the reason why Mele did not choose an example like this. If we rest satisfied with this point, we would already end up in a position that Mele's view has to give up a lot. For typical cases of *akrasia* do not involve our choosing against our judgment that something is "slightly better", or

²³ It is not clear what 'worse but permissible' could mean in this context. We can say that an option A is morally permissible, but worse than B, from a moral point of view. But this is tightly connected with the possibility of supererogatory actions. The concept of supererogation is difficult enough in the realm of ethics (for doubts about its applicability, see Shelly Kagan, "Does Consequentialism Demand Too Much?", *Philosophy and Public Affairs*, 13, 1984, pp. 239–54). It is even harder to understand how an action can be better all things considered yet supererogatory (all things considered?). To say that B is morally better than A, but that A is permissible, implies that doing A is not immoral. *Mutatis mutandis*, we would expect that the akratic action would *not* be irrational. In what sense, then, are we still speaking of *akrasia*? In what sense, can we then say that option A was better than B, all things considered, if it is not irrational to choose B? Perhaps, although choosing B is not irrational, it is more rational to choose A. But Mele surely owes an explanation of what it means to say that choosing A is more rational than choosing B, and yet it is not irrational to choose B. At any rate, since Mele gives no explanation about what it means to say that an option is permissible in this context, I will assume that it adds no content to the claim that an option is only slightly better than another.

choosing to do something that we consider to be in any interesting way “permissible”. In this case, Mele would have shown that only a small subset of akratic actions are free and this would already amount to giving up a lot of the phenomenon. However, the right answer is ‘yes’, but instead of making things better for Mele’s position, this fact just helps us illuminate why this reply to Watson is unsatisfactory even in the case Mele describes.

There is a clear reason why it is hard to imagine that Joe could fail to choose to exercise self-control and not change his judgment. For suppose he still makes this judgment, and he also judges that he cannot give in to his desire for cash without exercising this form of self-control. Of course, he might think that the exercise of self-control is painful enough that it changes his assessment of what is better or worse. If he thinks that a sign on the door is aesthetically so repulsive that it’s not worth posting it in order to save his sister’s life, then the exercise of self-control does not leave his judgment intact; reconsidering what is needed in order to save his sister, he no longer thinks that saving her life is better than accepting a million dollars. But as long as Joe is minimally logically competent,²⁴ and that he is otherwise indifferent about posting or not posting the sign on the door, he must conclude from the fact that it is better to save his sister’s life than to accept that money, that it is better to post the sign than not to post the sign. So we can safely conclude that, as long as Joe is not overcome by tremendous stupidity, he cannot maintain the original judgment that *A* is better than *B* (that saving his sister’s life is better than accepting the money) and also judge that it is better not to exercise self-control (that not posting the sign is better than posting it). However, he *can* fail to exercise self-control while still maintaining his original judgment. We can imagine that as Joe is about to post the sign he starts imagining all the wonderful cars he could drive, the wonderful places he could visit, the wonderful house he could buy if he had a million dollars—none of which, of course, compensates for the loss of his sister. We can imagine that despite not changing his reflective judgment he falls into temptation and does not post the sign.

In this case, however, Joe exhibits another instance of akratic behavior. Since Joe accepts the judgment that, all things considered, it is better to exercise self-control, if he doesn’t do it, he will be acting against his best judgment. And it should now be clear that the possibility of exercising self-control could not, in this case, help us explaining how the akratic agent was free. If the answer is “because he was free to exercise self-control”, it must be true that when the agent does act akratically, when he does not exercise self-control, he was still free to exercise self-control. But if his failure to exercise

²⁴ If he is not, we have a different kind of problem, similar to the case in which Joe does not realize that exercising self-control would let him act according to his best judgment. I’ll discuss this case below.

self-control itself involves *akrasia*, we cannot assume that he was free to exercise self-control unless we explain how *akrasia* can be free action.

The same analysis applies to the case Mele discusses, and we should not get confused by the fact that the agent judges the reason to be only “slightly better”. If he thinks that not going to the striptease show is better than going, and he is otherwise indifferent about using the desire-eradicating device, as long as he maintains his original judgment and is minimally logically competent, he must conclude that it is better to use the desire-eradicating device (given that he judges it to be a necessary, painless means to the option he judges to be better). Of course, he might still not use it, but this would be a case of going against his best judgment, since he must judge that all things considered, it is better to use the device. As we said above, however, if his failure to exercise self-control is due to *akrasia*, it cannot help us understand how akratic action can be free.

Mele also argues that the agent who failed to exercise self-control can still be considered free if she overlooked that it was possible to exercise self-control or if she misjudged the amount of self-control necessary, and thus did not realize that it was in her power to exercise self-control. So, for instance, suppose Joe thought that he did not need to post the sign on the door. Rather, he thought it would be enough to put on his sunglasses to avoid the temptation, or he just overlooked the possibility of posting the sign. Nonetheless, it is still true in this case that it was within Joe’s power “to resist acting on this desire”, and so “the desire did not *compel* him, in which case [his taking the money] was a free action.”²⁵

It is important to note, first, that on this account, again, only a small subset of cases of *akrasia* will turn out to be free actions. In most cases of *akrasia*, we cannot attribute failure to act on one’s best judgment to overlooking a certain technique for self-control or misjudging the amount of resistance necessary. Moreover, is it true that the akratic agent was not compelled by the desire in these cases? It might be useful to compare the possibility that Mele envisages with the case of someone who is held at gunpoint but does not realize that he can overpower his aggressor. It seems that this agent should not be considered as acting free from external compulsion, just because, unbeknownst to him, he could have successfully resisted his aggressor. In the same way, the agent who does not realize that she can exercise self-control should not be seen as acting free from internal compulsion.

Is an agent really free to do *X* if he can bring about *X*, but he does not know that he can bring it about? Am I free to escape from jail, if the door is open but I don’t realize that it is open? These are difficult questions to answer, but to accept Mele’s suggestion that the akratic agent was free to resist her desires we need to answer both affirmatively. Although I think the

²⁵ *Irrationality*, p. 22.

agent is not free in these cases, I do not wish to settle this issue here. I wish only to point out that even if there is a sense of ‘freedom’ that let us answer these questions affirmatively, Mele’s suggestion is not satisfactory. For, first, this sense of freedom does not allow us to say that the agent is *responsible* or *blameworthy* for acting akratically. At least we generally do not hold people responsible or blameworthy for not bringing about *X* if they are ignorant that they can bring about *X*.²⁶ If I could have saved someone’s life by flipping a switch, but I didn’t realize that I could have saved him this way I am not responsible for his death.

Secondly, *akrasia* in this case would not be irrational. For one cannot be accused of irrationality for not doing what she did not realize she could do.²⁷ So an akratic agent who overlooks the possibility of exercising self-control cannot be considered irrational on this account. But once she failed to exercise self-control, it was no longer up to her to follow her best judgment.

Finally, as Watson also points out, even if we think that the agent is free to act according to his best judgment in this case, what prevents the agent from acting according to his best judgment *while he is free to do so* is his ignorance, not *akrasia*. The failure of the agent, insofar as he is free, is ignorance (or misjudgment) not *akrasia*. Mele tries answering this objection as follows:

It may be true both that an agent would not have done *A* if he had seen what he needed to do to mount a successful resistance against his desire to do *A* and that, if he had been stronger or more resolute, a special effort of resistance would have been unnecessary.²⁸

Mele is suggesting that *akrasia* is at least one part of the explanation of why the agent did not do *A* when he was free to do *A*, since had he been strong-willed he would have done *A*. But the failure to do *A* *when* one was free to do *A* cannot be attributed to *akrasia* under this view. On this view, *akrasia*

²⁶ Unless, of course, the ignorance itself is culpable; if, for instance, the agent chose to engage in certain actions that would bring about his ignorance. In this case we would run to the same kinds of problem at a different level; we would have to explain why the agent decided to engage in these actions that cause him to be ignorant of the possibility of exercising self-control in order to bring about *X*. If the agent decided to engage in these actions, because he thought that the cost of knowing how to exercise self-control made bringing about *X* no longer an attractive option, the agent did not behave akratically—since, in this case, he thought that *all things considered*, *X* was *not* the best option. On the other hand, if the agent thought that all things considered, it was better to engage in these actions that would result in his knowing that he could (how he could) exercise self-control, but didn’t, then his failure to engage in these actions was an instance of *akrasia*. This can be considered blameworthy only if we have an account that makes *akrasia* blameworthy, but this is what is in question. For a recent discussion of the difficulties involved in the notion of “culpable ignorance”, see Michael Zimmerman, “Moral Responsibility and Ignorance”, *Ethics* 107, 1997, pp. 410–26.

²⁷ Again, unless not realizing that she could exercise self-control was itself irrational. And the same that was said in the note above applies here, *mutatis mutandis*.

²⁸ *Irrationality*, p. 27.

explains why there weren't other options open to the agent, not why the agent failed to choose an option that *was* open to him (or that he was free to pursue).

The problem of *akrasia* is *not* that we are *driven* by desires whose objects we do not value—this is merely compulsive action—but rather that we desire and positively evaluate that which we, under our own implicit or explicit recognition, ought not to so desire or evaluate. Thus, *akrasia* should be seen not as a collision between motivation and evaluation, but between two kinds of evaluation.

We have seen how to understand the conceptual tie between motivation and evaluation, and thus how to understand the scholastic formula. I argued that to desire something is to have a perspective from which it appears to be good (even if we reject this perspective upon reflection). We have found, however, that Davidson's attempt to provide an account of weakness of the will while accepting the scholastic formula was unsatisfactory. It left us without an understanding of how an agent could be prone to the kind of mistake typical of akratic behavior. We can clarify this point by taking a look at a suggestion by John McDowell. According to McDowell, the akratic agent makes the correct judgments about the good, but his grasp of the content of these judgments is defective. Let us look again at the akratic orthodox Jewish smoker. He judges, just as the virtuous person does, that all things considered, he should not smoke on Friday evenings; religious commands, after all, are more important than mere inclinations. What kind of grasp of the content of the judgment like 'All things considered, I should not smoke' could explain why the akratic agent stops short of acting in the appropriate manner?²⁹ What part of the phrase 'all things considered' did the agent fail to understand? What we need to explain is how the structure of our practical deliberation makes us prone to form unconditional judgments that disregard our most considered, reflective judgments—a kind of mistake that on the face of it seems incomprehensible.

We have also seen why the rejection of the scholastic formula leaves us without a satisfactory account of *akrasia*. Now, if we think that *akrasia* is a phenomenon which has no parallel in the realm of theoretical reason, it is natural to look at unique features of practical reason to account for *akrasia*. Thus, it might be tempting to look for the source of *akrasia* in a sort of motivational breakdown, because there is no similar form of breakdown in

²⁹ I am also changing McDowell's terminology to make it consistent with the rest of the paper. See his "Comments on Irwin's 'Some Rational Aspects of Incontinence'" *Southern Journal of Philosophy*, 1988, suppl., pp. 89–102; esp. section V. Of course, McDowell also tries to explain the possibility of a defective grasp of the content, but it is unclear that McDowell's characterization of the failure of the akratic agent is compatible with a scholastic view. See, for instance, his claim that "The incontinent's problem is rather that practical thinking with this sort of content is not (perhaps not yet) fully ingrained into his motivational makeup" (p. 97).

theoretical reason. In the next section I want to examine Descartes's views on the relationship between the will and the intellect, and how they show us that there is a theoretical counterpart to the phenomenon of *akrasia*. I will also argue that we can find in Descartes's work the resources to account for this possibility. In the following section I will show how to extend this account to practical reason—to the more often discussed cases of *akrasia*.

But before we move on, I must clear up a possible source of misunderstanding. One might think that by adopting Descartes's view of the relation between the will and the intellect, I have adopted a highly controversial form of voluntarism: the view that we can change our beliefs as we please, or that we can decide to believe whatever proposition we want to believe. Or in a less extreme version that in at least some significant cases, we can come to believe a proposition solely by wanting to believe it. Although some people might want to defend some version of this kind of voluntarism,³⁰ I do not think Descartes is committed to it.³¹ Descartes does think that affirming, denying and doubting are modes of the will, and so that believing involves the will. But this identifies believing with a form of activity, and in particular free activity, but not necessarily with the ability of believing whatever one wants to believe.

Although the nature of this activity will be clearer below, it might be useful to put forth a word of warning against a confusion that could tempt us to ascribe to Descartes this controversial form of voluntarism. If we take judgment to be the main activity of the will, we can distinguish between judging to be true and judging to be good. From what we said above, it should be clear that it is natural to identify that which we judge to be good with that which we want. So it would be natural to equate our freedom to act in accordance with what we judge to be good—a capacity of self-determination—as a freedom 'to do whatever we want'. But it would be misleading to think that we must be able to characterize our freedom to act in accordance with what we judge to be *true* also as freedom 'to do whatever we want', since judging to be true is not conceptually connected to wanting. If we want to draw a parallel with this expression in the realm of theoretical freedom we must say that freedom to act in accordance to what we judge to be true is our freedom to accept whatever seems correct to us, not that this freedom is the freedom to believe whatever we please.³²

³⁰ As far as I know, no one claims that, for any proposition, no matter how absurd it is, one can affirm it or deny it at will. For some (less extreme) versions of voluntarism, see Barbara Winters, "Willing to Believe" *Journal of Philosophy*, LXXVI, 1979, pp. 225–43 and Bas van Fraassen, "Belief and the Will" *Journal of Philosophy*, LXXXI, 1984, pp. 235–56.

³¹ For a different view, see Harry Frankfurt, *Demons, Dreamers and Madmen* (Indianapolis: Bobbs-Merrill, 1970).

³² One might think that this way of defending Descartes from a controversial form of voluntarism could not account for passages in which he characterizes our freedom as

III

The possibility of theoretical *akrasia* comes up most clearly in Descartes's *Meditations* when we try to carry the skeptical lessons of the First *Meditation* into the later *Meditations*. At the end of the First *Meditation*, Descartes expresses the problem as follows:

My habitual opinions keep coming back, and despite my wishes, they capture my belief, which is as if it were bound over to them as a result of long occupation and the law of custom.³³

This is not very far from how an akratic agent could describe his failed attempts to follow his reflective judgments. But what makes those "habitual opinions" come back, flouting the skeptical arguments of the First Meditation? Why can't his "wish" be easily fulfilled in light of the doubts raised against his habitual opinions? In other words, why aren't the skeptical arguments enough to make sure that old habits will not capture his belief?

I think the answers to these questions can be better understood when we look at a distinction implicit in Descartes's reexamination of these doubts in the Third Meditation. When Descartes conceives the possibility of radical doubt, he notes that whenever he contemplates thoughts such as ' $2 + 3 = 5$ ' he cannot fail to assent to them. As he says:

when I turn to the things themselves which I think I perceive very clearly, I am so convinced by them that I spontaneously declare: let whoever can do so deceive me; he will never bring it about that I am nothing so long as I continue to think I am something; (...) or bring it about that two and three added together are more or less than five, or anything of this kind in which I see a manifest contradiction.³⁴

The possibility of skeptical doubt depends upon the possibility of not "turning to the things themselves"; that is, not turning my attention to those truths which I cannot doubt when I contemplate them, while still retaining them in the mind somehow. I take it that Descartes is here drawing a distinction between oblique and direct thought. There is a way in which I conceive

"freedom of indifference". The exercise of this kind of freedom does not seem to depend on judging something to be true or good. But, since affirming and denying are always forms of judging, the difference between freedom of indifference and freedom considered merely as spontaneity is a difference between the *grounds* of our judgments of the good and true, not a difference about whether we make judgments about the true or good (since our freedom is always expressed in the activity of judging). I believe that the clearest discussion of these issues in Descartes's work is to be found in a letter to Mesland in *The Philosophical Writings of Descartes*, vol. III, trans. by Cottingham, J., Stoothoff, R., Murdoch, D., and Kenny, A. (New York: Cambridge University Press, 1991, henceforth "CSMK"), pp. 244–46, and in *Oeuvres de Descartes* (henceforth "AT"), edited by Ch. Adam and P. Tannery (Paris: Vrin/C.N.R.S., 1964–76), vol. IV, pp. 173–75.

³³ *The Philosophical Writings of Descartes*, trans. by Cottingham, J., Stoothoff, R., and Murdoch, D., (New York: Cambridge University Press, 1987; henceforth "CSM"), vol. II, p. 15; AT VII, 22.

³⁴ CSM II 25; AT VII, 36.

of two plus three equaling five such that this thought presents itself to me as impervious to doubt, a thought in which I grasp not only *that* two plus three equals five but *how* it is so. I will call the former an oblique cognition and the latter a direct cognition.³⁵

We can say, roughly, a direct cognition is a representation of an object or claim through which one clearly understands (or seems to understand) *why* the object is as one represents it, or *how* it is that this claim is true.³⁶ An oblique cognition is a representation of a claim or an object which is not a direct cognition, but one through which one understands (or seems to understand) *that* there are reasons to accept that the object is as one represents it, or that the claim is true. We can think of an oblique cognition as standing proxy to one or more direct cognitions. The direct cognitions carries with it the explicit justification of that which the cognition represents as true. So, a mathematical proof of a theorem is a direct cognition of this theorem, since it not only presents the theorem as true, but also presents the reasons for holding this theorem to be true. But if I later remember having proven this theorem to be true, without remembering the proof itself, my cognition is oblique. It represents not *the* reason for holding the theorem to be true, but that *there is* a reason to hold this theorem to be true, a reason that this cognition does not make available to me. Of course, there is a sense in which my memory is itself a reason for my accepting the truth of the theorem, but it is a reason for *accepting* the theorem, not a reason *why the theorem is true*; my memory could not be part of the proof of the theorem. In the same way, if I trust the judgment of reputable physicists on matters of quantum mechanics, my cognitions of the matter are oblique; they stand proxy not to a direct cognition that *I* have ever had available to me, but they nonetheless stand proxy to a direct cognition.

In both direct and oblique cognitions we deal with the same proposition, but we conceive of this proposition in different ways. We can also see that this distinction is important in motivating Descartes's intuitionism. Descartes insists that we should try, as much as we can, to entertain several steps of a derivation at once.³⁷ This insistence can be seen as manifesting not a bizarre mistrust in the powers of memory but rather a realization that this is what true understanding partly consists in. The mathematician who does not see the whole proof of a proposition within a single intuition knows *that* it

³⁵ I do not intend to use the word 'cognition' here as an achievement noun. So it does not follow from the fact that I have a cognition that presents me an object as *X* on the grounds that *Y* that the object is in fact *X* or that *Y* constitutes good grounds for accepting that the object is *X*. My cognition might be in some way delusory.

³⁶ This does not mean to say that one can say anything more than "I see it!" to justify why the object is as she represents it.

³⁷ See Rule 11 of "Rules for the Direction of the Mind" (CSM I, 37; AT X, 407).

is so and even *that it must be so* but does not know *how* it is so and *how it must be so*.

Oblique cognitions seem to be especially important in reflective judgments, since reflective judgments might have to weigh considerations that are not clearly commensurate. By looking at two objects at more or less the same distance, I can determine, in most cases, which object is larger just by looking at them. But the situation is more complex, if I have to examine, for instance, incompatible claims of my sensory cognitions of an object and my more theoretical understanding of a physical object. Suppose, for instance, I see what seems to be an object floating in the air in flagrant violation of the laws of gravitation. When one asks whether the claims of our theoretical understanding of physical objects should be valid or whether the claims of sensory perceptions should be valid, one is not just struck by each of these cognitions and put in the position of waiting to see which of them would incline the will one way or the other, or by just bringing these different ideas to one mind. I have to consider how I should take each idea, when and how I should take my sensory perceptions at face value, the reliability of my clear and distinct perception (or my theoretical beliefs about the nature of the object), and so forth. In order to settle the competing claims of those cognitions, we need to rely on oblique cognitions, on cognitions such as: 'the fact that I have a sensory cognition of the state of a physical object is a reliable but not infallible guide to the state of this object (it should count as some evidence to the claim that the object is in this state, but not as conclusive evidence)'. In cases of conflict of cognitions from different cognitive sources, our reasoning involves not only the presentation of these cognitions to the mind, but primarily the representation of the *fact that we are in the possession of these cognitions*, and of our reflective views on the relative weight that we should accord to these cognitions. We can now reserve the term 'reflective cognition' only to those cognitions that evaluate the relative weight of a certain cognition, such as the following: 'the fact that I have a sensory cognition of the state of a physical object is a reliable but not infallible guide to the state of this object (it should count as some evidence to the claim that the object is in this state, but not as conclusive evidence)'. I will call a 'primary cognition' any cognition that is not reflective. All reflective cognitions are oblique, but not all oblique cognitions are reflective. We can call a 'reflective judgment' a judgment in which one attempts to settle the incompatible claims of different sources of cognition. A reflective judgment will be based on reflective cognition, and thus on oblique cognitions. Note that this is not a claim about empirical psychology or phenomenology, but a claim about what must be involved in our capacity to adjudicate the claims of incompatible cognitions. Since each cognition presents the object as corresponding to this cognition, we cannot settle this question through, so to

speak, the eyes of the cognitions themselves, but only as we reflect upon the significance of having these cognitions.

We are now in a better position to understand Descartes's separation of the will from the intellect in a way that does not commit Descartes to viewing the will as an arbitrary capacity to pick and choose its favorite ideas as objects of belief. It has often been pointed out that Descartes thinks that I cannot fail to assent to the fact that a triangle has three sides.³⁸ Why does Descartes then insist on separating the will from the intellect, in distinguishing my assent to the idea from my understanding of it? One could limit the activity of the will to the cases when we do *not* have a clear and distinct idea, but this would fly in the face of Descartes's assertion that we are most free when we assert that which we clearly and distinctly perceive.³⁹

This separation would be indeed unnecessary had we no ideas that, in an important sense, compete—ideas that are alternative presentations or apparent presentations of the same object or fact. The idea that we have of a corporeal object when we are walking around seeing corporeal bodies is very different from the one that we get from learning physics. The first presents it as colored and continuously solid in a way that the latter does not. Descartes's own example is the two ideas we may form of the sun, one from sensory perception and the other from physics. But it is important to conceive them as different appearances of the same thing, so that they can be seen as delivering two incompatible views of the same object. Moreover, we could not understand them as presentations *of the sun* if in some way they did not appear to us as (accurate) representations of the sun; or, in Descartes's words, as if they were images of things.⁴⁰

However, unless we want to think about a subject in rather schizophrenic terms, we have to make room for the fact that the subject, at least in most cases, adjudicates between competing representations—a person whose thoughts present incompatible views of the same objects has to decide which thought is the one that she should accept and which she should see as merely an *appearance* of a certain content being affirmed rather than a true affirmation of the content. Without a concept of the will, we have to conceive of the subject as thinking that the sun is smaller than the earth in his daily business and as having a totally different idea of the sun when he is engaged in science or reflective thought. If it were not for the fact that the subject makes *judgments*—i.e., the fact that a subject affirms some ideas and denies others—the competing claims of these ideas would simply coexist. In order to be disabused of the daily illusions the subject must be capable of bringing

³⁸ See, for instance, John Cottingham "Spinoza's critique of Descartes", *Journal of the History of Philosophy* 26:2, 1988. But see letter to Mesland (CSMK, 244–46; AT IV, 173–75) for an important qualification of this claim.

³⁹ See CSM II, 40; AT VII 57–58.

⁴⁰ CSM II, 25; AT VII, 37.

these two ideas together and of taking one to be that which provides her with the correct representation of the sun. The activity of the will is then this activity of adjudication. Contrary to what one might think, the will turns out to be a truly cognitive faculty. In a sense it is the cognitive faculty *par excellence* since it includes the capacity to weigh the different claims of different ideas of the same object of cognition.

Under this picture we get a conception of the will as evaluating competing representations of an object; the will turns out to be a faculty of the agent closely connected to our ability to consider him rational. It is because the will can adjudicate between the claims made by different ideas of the same object that it can be considered a rational capacity, through which inconsistency and incoherence can be avoided. One can think about the Cartesian will as performing two different tasks. The first is to bring to the gaze of the mind primary cognitions of ideas of the same object. The will has to be able to represent the claim made by, for instance, the idea of the wax given by sensory perception, and the one made by the intellect. But the will, of course, is not merely a reflective capacity but an adjudicating one also. If it is performing its job correctly, it will affirm only the clear and distinct perceptions of the intellect (at least when we are engaged in the search for truth). In the case of intuitive judgment this could not fail. The clarity and distinctness of the primary cognition suffice to guarantee assent. But if, as a result of an abstruse derivation, the will has to rely upon oblique cognitions, the will might end up making the wrong judgment, given that the reflective cognition will inherit the obliqueness of the primary one.

If, when going through the reasoning of the First Meditation all that I carry from it is the thought *that* I had reasons for doubting the existence of the external world, I might not withhold my judgment accordingly (though I might accept that I ought to withhold judgment). If my will compares this mere recollection with the ideas of sensory perception—which provide me, at least apparently, with primary cognitions that represent this world—I might find the latter convincing while still agreeing that I ought to be convinced by the former (and this might in practice be the fate of every skeptic). If this happens, we have a case of theoretical *akrasia*. This brings us back to a point I made earlier. The more clearly and distinctly I understand the grounds for a certain truth, the less am I prone to this sort of theoretical *akrasia*. Descartes's advice in Rule Eleven that we should keep as much of a proof in mind as possible is all the more important if we realize that this clear understanding of a proposition can also prevent us from slipping into this form of theoretical *akrasia*. At any rate, it is important to note that the fate of the skeptic is not unlike the fate of the akratic agent. In the same way, we can say that the orthodox Jewish smoker thinks that he ought to be convinced by his reflective judgment, but he is actually convinced by the judgment from the point of view of pleasure.

This distinction between what I believe and judge as opposed to what I believe I ought to believe and judge, should not be seen as a distinction between the way states of belief strike a helpless agent, and the way the agent would like to find himself believing. *Both kinds of beliefs express (freely made) judgments.* We should think of this case as the case of someone who has been convinced by a philosophical argument, but who, when turning her attention from the argument, cannot fail to be persuaded by the evidence of her senses. It is no surprise then that we resort to a device such as the evil demon to make this clear. In this way the upshot of the First Meditation can be surmised in a direct cognition. And here again we can see that the evil demon has a function similar to the various devices that help us exercise self-control. The evil demon performs the same function for the Meditator that the conspicuously displayed autopsy photos of the lungs of a heavy smoker perform for the person trying to quit smoking. And here too perhaps we can say that his reflective judgment is based on a reflective cognition. As such, no matter how certain it is, its certainty is grounded on oblique cognitions. And if this is the case, the akratic agent, just like the skeptic, might end up adopting the unreflective point of view of her direct cognitions.

It is important to distinguish what I have been calling theoretical *akrasia* from a much discussed phenomenon: self-deception. Mary gives Joe all the indications, short of just saying it, that she considers him a bore. She yawns while he speaks, she avoids his company at parties, and she makes no effort to feign interest in his anecdotes. Joe thinks very highly of Mary's judgment and it would be devastating to him to believe that she finds him a bore; he certainly wants to go on believing that she finds his company invaluable. Despite all evidence, Joe believes that Mary does not consider him a bore.

We can make the distinction clear by pointing out that theoretical *akrasia*, as I understand it, is a shortcoming wholly within theoretical reason. Its description involves no reference to the desires, values or practical judgments of the agent. Self-deception involves the interference of practical reason; Joe's desires are determining the formation of an irrational belief. One is not motivated in any way to find the gambler's fallacy persuasive; it would be quite surprising to find out that one has a stake at that issue. Yet, one finds it persuasive even if one acknowledges that one ought not to find it persuasive.

Depending on how one understands self-deception, it will be a particular case of theoretical *akrasia* or a distinct phenomenon. If one thinks that the self-deceived agent accepts that he ought to form the contrary belief (if one thinks, for instance, that in order to be self-deceived Joe must accept that he ought to believe that Mary considers him a bore), then self-deception will be a particular case of theoretical *akrasia*. If one thinks, perhaps more plausibly, that self-deception does not involve accepting that one should form the contrary belief, then it will turn out to be a phenomenon distinct from theoretical *akrasia*. In either case self-deception will raise difficult and interesting issues,

but these issues are not the topic of this paper. My claim is that there is a phenomenon analogous to practical *akrasia* that can be described wholly within the realm of theoretical reason that can help us illuminate its better known practical counterpart.⁴¹

IV

We can now see how this pair of distinctions—oblique and direct cognition; primary and reflective cognition—can help us provide a scholastic account of weakness of the will. We can think about practical reason in a similar way. We can think that what makes our practical sensibilities cognitive capacities is that we can see them as primary cognitions whose claims are evaluated in reflective cognitions in our exercise of practical reason. This is to say that a desire, an urge or any other form of motivation provides us with a putative perspective of the good, which we may call a ‘cognition of the good’ to be evaluated upon reflection. Under this understanding, an urge which is taken into consideration when deciding how to act or how to lead one’s life, should not be seen as a mere impulse but as a putative judgment, a way of seeing the worth of an action. Comparative judgments in which we try to settle the incompatible claims of different desires can be seen as reflective judgments, in which we weigh the force of the claims made by competing cognitions of the good.

Any judgment, for instance, which has to settle between claims to the good from different perspectives presupposes a comparison of oblique cognitions of the competing cognitions of the good, which are not immediately commensurable.⁴² Since different desires might present an object as good from different perspectives they cannot be evaluated through direct cognitions;

⁴¹ For the same reason, what I take to be theoretical *akrasia* is distinct from what Mele takes to be akratic belief, since he claims that “‘motivatedness’ of incontinent action is one of its defining features. This is true as well of incontinent believing” (*Irrationality*, p. 112).

⁴² By saying that two kinds of goods are not commensurable, I do not mean to say that there can be no reasonable grounds to choose between them. All that is meant is that we cannot find a common measure that would allow for a weak ordering of all goods of these kinds. Suppose I am deciding between how to spend some extra \$1000 I have received. I could either invest the money in my retirement plan (R) or use it to contribute to the election of a certain political candidate (E). Suppose I am indifferent (or find that I can’t decide) between the two alternatives. Suppose now that my retirement plan announces a promotion that will, upon my request, add a \$10 bonus to my \$1000 if I invest in the plan (R*). It is clear that I prefer R* over R even if I might still be indifferent (or unable to decide) between R* and E. So we get that $(R \geq E)$, $(E \geq R^*)$, and $\neg(R \geq R^*)$. For a similar notion of incommensurability, see Joseph Raz, *The Morality of Freedom* (New York: Oxford University Press, 1986), p. 232. See also on this issue Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, Mass: Harvard University Press, 1993), ch. 3. For an account of *akrasia* that relies heavily on a notion of incommensurable modes of valuation, see, David Wiggins, “Weakness of Will, Commensurability and the Objects of Deliberation”, in A. Rorty, *Essays on Aristotle’s Ethics*; Berkeley: University of California Press, 1980).

thus, the ensuing reflective cognitions will inherit the obliqueness of the primary cognitions upon which they are based. When I recall the thrill of watching a mystery movie, and I imagine myself reliving this experience, I have a cognition in which watching mystery movies is presented as desirable. However, when evaluating what should I do tonight, I have to weigh the claim of this cognition against other incompatible claims. When reflecting about this option, I weigh that watching a movie is desirable because of this thrill, but in doing so I do not necessarily see *how* the thrill makes it desirable; I am no longer judging what is good from the perspective of desires for thrills. My reflective judgment here too is based on reflective, oblique, cognitions. The reflective judgment will be grounded on cognitions such as: ‘the fact that I find *X* pleasant is a reliable but not infallible guide to the fact that *X* makes some positive contribution to my good’. As I recognize my reflective judgments as authoritative, I recognize that my unconditional judgments *should* always come from my reflective perspective. But unless the reflective cognition can do away with the natural plausibility of the other perspectives, here, just as in the case of theoretical *akrasia*, the less clear my grasp is on the reflective cognitions, the more prone I will be to form unconditional judgments that do not *in fact* arise from my reflective perspective. Let us now examine this claim in more detail.

Let us return to our orthodox smoker. He judges that:

- (i) From a certain point-of-view, smoking is better than refraining from smoking.
- (ii) All things considered, refraining from smoking is better than smoking.

As we said above, Davidson correctly points out that (ii) is compatible with the unconditional judgment that it is better to smoke. But we argued that though Davidson’s account of *akrasia* could make these statements compatible, it could not explain the frequency of instances of their conjunction in free action.

The root of this difficulty lies in Davidson’s account being oblivious to the fact that when we move from judgments of the form of (i) to judgments of the form of (ii), something gets lost. In order to make judgments of the form of (ii), we have to step back from the perspectives under which judgments of the form of (i) are made. If we lose sight of this, we will be bound to think that (ii) makes (i) irrelevant—at this point (i) could have no bearing on his action if the agent can follow his judgment. After all, (ii) seems to do away with any plausibility that (i) might have. But we can understand what gets lost in this move in light of the pair of distinctions we brought forward in our discussion of Descartes.

A desire provides the agent with a direct and primary cognition. An object of a desire appears in this way to be good and worthy of being pursued. When reflecting upon such a desire, one notices that its realization is incompatible with more important ends. So the end of eating chocolate cake might be incompatible with keeping oneself healthy. Or a cognition of the good is deemed illusory upon reflection. For instance, Walter might see his desire to avenge himself against his rival who got the poetry prize he so much coveted as arising from ungrounded emotions. After all, to the extent that desire for revenge makes sense, it ought to be directed against those who wronged him, those whom he might legitimately resent. But Walter cannot take seriously the thought that someone who got a poetry prize he coveted indeed has wronged him. So these reflective cognitions present him with the view *that* the primary cognition is illusory or overridden. But the primary cognition remains, and the reflective cognition might not have the same immediacy or clarity that the first cognition has.

So in the same way that convincing myself that the sun is bigger than the earth does not thereby do away with the way the sun looks to me when I step out of my house, the judgment that he ought not to try to avenge himself does not do away with the original cognition of the good. A certain cognition of the good that I believe I ought to accept might fall short of immediately providing me with an all-out judgment in at least two ways. I might understand that I should pursue a certain course of action, but have very little insight as to why. This could happen if I were told by someone I trust that a certain course of life is worth pursuing (she explains to me, for instance, how one feels much better when one stops eating animal products) or if I were convinced by complex considerations that I ought to become a vegetarian. But it might also fall short of immediately providing me with an all-out judgment if I can see why I ought to accept this cognition of the good, but I cannot clearly grasp the alternative cognitions as illusory.

This is similar to a situation in which a *reductio* makes it clear to us that a certain proposition cannot be true, but we do not understand what is wrong with the reasons we took to support this proposition. Again, we can think about the fate of skepticism at the end of the First Meditation in the same terms. The remembrance of having good reasons to doubt all of one's beliefs tells us that that which our senses convey to us ought to be doubted. But the conviction that this was shown is an oblique cognition whereas the presentation of objects by the senses is immediately clear. And if my understanding of this argument is unclear and vague I might be unpersuaded by such an argument while still retaining my conviction that I ought to be persuaded. Similarly, realizing that the gambler's fallacy is a poor form of reasoning does not necessarily dissolve the temptation to avoid betting on last week's winning number, and this temptation is greater the less clear my grasp of the structure of the fallacy is. So I will probably be differently persuaded by

instances of the gambler's fallacy if I was just assured by a friend that this is a fallacy, or if I have merely a vague understanding of how the principle guiding my choice is an instance of the gambler's fallacy. And if this is the case, I might find myself saying things such as: "I understand that this is a fallacious reasoning, but it just does not seem possible that these numbers will come up again".⁴³

Turning back to the akratic person we may say, similarly, that the akratic agent's understanding that A is better than B is a reflective cognition of which she might have only a vague or oblique understanding. And in the same way that someone who has read the First Meditation might believe that he ought to be persuaded by the argument, while not actually being persuaded by it, the akratic agent believes that he ought to be persuaded by his reflective understanding, but he is not persuaded by it. And this amounts to saying that the akratic agent acts against his own recognition that all things considered he should act otherwise.

When Walter vividly imagines that person holding with pride the award he so much coveted, he can see the *point* of hating that person and taking revenge. Walter knows indeed *that* he ought not to take revenge, but his desire for revenge might present a conception of the good whose inadequacy he only vaguely understands right now. Walter might see *why* he should take revenge in a way that he cannot see why he should refrain from doing so. Or Walter might see *that* he should not take revenge, but he cannot clearly see *why* he shouldn't—the plausibility of the desire for revenge is still there. Of course, Walter could, so to speak, wise up. His understanding of the grounds for the inadequacy of his desire for revenge could become more and more clear to him, in the same way that it will require some effort for a mature human being to acknowledge that, in a sense, the stars in the sky 'look small'; a mature human being will often just perceive the stars as distant. In the ideal case, Walter will become virtuous in the Aristotelian sense of 'virtuous',⁴⁴ and he will no longer be prone to akratic behavior. His understanding that there is nothing to be said for this perspective will become flawless, and his desire for revenge will be lost. Walter would still understand why one might believe that there is something to be said for revenge in these situations, how

⁴³ One might think that if a person does not bet consistently with the belief that these are instances of the gambler's fallacy, then, at least when it comes to the time of betting, he does not have available (at least at that moment) the understanding that this is a fallacious form of reasoning. But the same goes for practical *akrasia*. One can always claim that the agent who acts against her best judgment no longer has available her reflective understanding that, say, A is better than B. There is no more reason (at least no more pre-theoretical reason) to deny the phenomenon in one case than in the other.

⁴⁴ The virtue in question, for Aristotle, is temperance. According to Aristotle, "the continent and the temperate person are both the sort to do nothing in conflict with reason (...); but the continent person has base appetites, and the temperate person lacks them." (*Nicomachean Ethics*, 1151b–1152a)

this understanding of what can count as good could make a claim on *someone*, but it would no longer make a claim on him. Walter would not see revenge as good from any perspective, and so, according to our scholastic view, desire for revenge would not in this case be one of his desires. What the virtuous agent accomplishes is that his unconditional judgment no longer needs to rely on oblique cognitions, since there are no longer any direct cognitions of the good that present to him a temptation as a good worth pursuing in this occasion. Since those things that could be objects of temptation do not appear to be good to this ideally virtuous agent, the reflective perspective is idle: it does not need to compare or weigh the claims from different perspectives.

Perfect virtue is certainly not easy, and likely impossible. It is no surprise that we devise methods that are similar to Descartes's evil genius: a direct cognition that can mimic the function of perfect understanding. One might look at a lovely picture of one's family to stay away from a gambling table, or keep large sized clothes that one used to wear in order not to indulge again in old eating habits. Of course these strategies are just the same strategies that come under the heading of self-control and that are invoked by those who ascribe *akrasia* to a gap between motivation and evaluation. Self-control would be a way to line up one's motivation with one's evaluation. But on the view I am defending self-control should be understood as a way of getting one's evaluation attuned with one's conception of how one ought to evaluate. It is not a process of garnering motivation, but finding a clear and obvious way to present something that one abstractly and perhaps vaguely judges to be good—in the ideal case, finding a direct cognition that can substitute for an oblique one. That this is also a better account of how we generally conceive of these strategies can be seen by comparing the examples above with, for instance, a pill that would create an irresistible craving to do the right thing. It is important to imagine that the pill does not make it vivid to me how much better it is to do the right thing, but just creates in me what one could describe as a "surd" form of craving for, for instance, saying 'no' to any form of bribery, or a surd form of revulsion to accepting bribery.⁴⁵ If we accept the account of self-control according to which self-control aims at garnering further motivation that will match a fixed evaluation of the options, taking the pill would be a form of self-control like any other; it does indeed garner motivation that will make me act according to my evaluation. Undoubtedly, given my tendencies to *akrasia*, I could decide to take the pill. However, it would be odd to describe this as an exercise of self-control. Quite the contrary, it would seem a case in which I despair of exercising self-control

⁴⁵ A classical fictional example of this sort of revulsion is the revulsion developed by the central character of "Clockwork Orange" towards violent crimes.

and let myself be dragged by my passions, and, in light of this fact, try to make sure that they will at least drag me in the right direction.

Although perfect virtue might be an unattainable ideal, a growing understanding of the grounds for a reflective judgment might end up doing away with the competing *prima facie* value judgments. At some point, one learns that some jokes are in bad taste, that, for instance, one is not supposed to make witty remarks about someone else's distress, or to laugh if someone makes this kind of remark. At first, it might be hard to see the point of this; it might require some argumentation to convince a naughty child not to engage in this kind of behavior. If, on this occasion, the child hears a joke in bad taste she might still laugh, or it might take her some effort to control her laughter. As it becomes more and more obvious to her that one should not make such jokes, the situation entirely changes. And the change can be seen not only in the fact that she never laughs at these jokes, that there are no more relapses. We can see the change also in the fact that she will no longer find these jokes funny at all.

The possibility of this virtuous agent makes it clear that according to this account of *akrasia*, akratic behavior always involves a form of cognitive failure. This might be seen as a denial of the phenomenon, especially because it might look as if it were incompatible with a judgment of the form of (i). But this appearance is due to the assumption that all persons are cognitively equal with respect to the following claim (R) as long as they assent to it:

(R) All things considered (from a reflective perspective), A is better than B.

However, there are quite different ways in which one can be said to know (or believe) that (R) is true. In the case of the virtuous agent, she knows not only that (R) is true, but why (R) is true and how (R) silences the claims of alternative conceptions of the good. And thus her knowledge does away with any illusory appearance of good that B may have due to a certain desire.

It is important to note, however, what sort of cognitive failure this is, so we can see that, in an important sense, it does not rule out the possibility of "clear-eyed *akrasia*". Consider an akratic agent and a virtuous agent who is not prone to *akrasia*. Let us assume that they make the same reflective, or all things considered, judgment that A is better than B. The difference between the akratic agent and the virtuous agent is not a difference about the *content* of their belief of what is best, all things considered. They both accept that A is better than B all things considered. Also they both agree, as broadly rational agents, that one ought always to follow the reflective perspective. So the akratic agent believes that:

- (a) From a reflective perspective (all things considered), A is better than B
- (b) I ought to accept the judgments from a reflective perspective.

Knowledge of (a) and (b) should certainly suffice, insofar as the agent is rational, to issue in the same all-out judgment as the virtuous agent, that is:

- (c) A is better than B.

Indeed, as one would expect from any account of *akrasia* that does *not* deny the phenomenon, the puzzle is why the agent, insofar as she is still a free agent, does not act exactly like the virtuous agent. Insofar as one holds a scholastic view, this puzzle will come together with the puzzle of why the akratic agent, unlike the virtuous agent, does not endorse (c),⁴⁶ and, of course, this paper has been aiming at dispelling this puzzle.

It is important to note that this account preserves the truth of most of the statements that we would make to describe an akratic agent. It is true, under my account, that the akratic agent acts against her best judgment (her reflective judgment), and that she did what she knew she should not have done (she knew she should have accepted the all-out judgment that corresponded to her reflective judgment), and she acted in a way she knew to be irrational (she accepted the all-out judgment she knew she should not have accepted). But does it preserve *everything* we ordinarily say about weakness of the will? Does it preserve the simple judgment that the akratic agent chose B, even though she judged that A was better than B? It depends on whether by “judging that A is better than B” we mean the all things considered judgment or the all-out judgment. My account, of course, can accommodate only the former. The question is whether we are required to use the latter interpretation to save the phenomenon. The fact that we can preserve all the above ordinary judgments using the former interpretation suggests that the answer is ‘no’. The former interpretation still allows us to describe the agent as freely choosing that which she knows she ought not to choose, and this is the essence of the puzzling phenomenon of *akrasia*.⁴⁷

⁴⁶ The same is true, for instance, of Davidson. On his account, the akratic agent always accepts the all-out judgment that corresponds to the action. See “How is Weakness of the Will Possible?”.

⁴⁷ Charlton points out, in criticizing Davidson’s account of *akrasia*, that the akratic agent will not agree with Davidson’s characterization of her. The akratic agent will deny making the unconditional judgment that the course of action he undertook was the best one (*Weakness of the Will*, p. 124. Michael Bratman makes a similar point in “Practical Reasoning and Weakness of the Will”, *Nous* 13, 1979, pp. 153–71). However, the distinction between an unconditional judgment and an all things considered judgment is part of the philosophical lore. It is not a distinction that one is ordinarily aware of, but rather depends on a particular theory about the nature of practical reason. The fact that an

Of course, it might be the case that on a certain philosophical account of the phenomenon, one needs to accept the latter interpretation. It is important not to confuse the phenomenon and its putative accounts, hard as it is to keep them apart. One cannot demand that a philosophical account will preserve all the features attribute to it by putative philosophical accounts. One should keep this in mind when considering whether, under this account, clear-eyed *akrasia* is really possible. For one might say that there is, after all, a cognitive difference between the akratic agent and the virtuous agent. So, isn't only the virtuous agent really clear-eyed?⁴⁸

But, as I said above, the difference is not a difference about their acceptance of the *content* of the reflective judgment, but rather their understanding of its *grounds*. So the akratic agent knows all there is to know in order to choose the right course of action. The akratic agent *knows exactly what he ought to do*, but doesn't do it; this is an appropriate description of the phenomenon that my account preserves. Indeed the lesser grasp of the grounds of the reflective judgment of the akratic agent, as compared to the virtuous agent, does not, by itself, make an agent akratic—it just makes him *prone to akrasia*. The strong-willed agent does not succumb to this temptation.⁴⁹

One can think of the akratic agent as someone who has a desire whose presentation of an object wins over his reflective understanding of the good, as someone who is so persuaded by the desire that he no longer accepts that the reflective understanding presents the better reasons, though he accepts that he ought to think otherwise. The phrase 'win over' should not mislead us to think that this happens despite the agent. It is the agent who fails to be persuaded and it is her own judgment that these reasons are overriding, just as it is the would-be skeptic's own judgment that the external world is still there.

One can explain the difference between my view and the view that sought to sever the ties between motivation and evaluation as follows: Those views correctly diagnosed the problem of *akrasia* as a gap between the way in which I think things ought to happen and a way in which they actually happen. But

agent claims that she does not consider the action she undertakes better than the one she does not cannot settle whether she is expressing her all things considered or unconditional judgment. Insisting that we interpret it as the latter is already the result of adopting a certain philosophical position on the matter.

⁴⁸ Of course I am not the first one to propose that understanding the akratic agent in a way that involves some cognitive failure still preserves the phenomenon. Aristotelian accounts of *akrasia* will typically attribute to the agent some cognitive failure. Norman Dahl, for instance, tries to show that "Aristotle did allow genuine cases of weakness of the will," (p. 140) while providing an interpretation of his account of *akrasia* that "a person cannot act contrary to full and complete knowledge of how one ought to behave" (p. 189). See *Practical Reason, Aristotle and Weakness of the Will* (Minneapolis: University of Minnesota Press, 1984), esp. Part II. See also McDowell's "Comments on Irwin".

⁴⁹ One may ask why then the akratic agent does not behave in the same way as the strong-willed agent does. In the last section of the paper, I explain why I think there is no philosophically interesting answer to this question.

they took it that once this gap opens it cannot be a gap fully described in cognitive terms, terms which involve an essential reference to judgments made and endorsed by the agent. As long as all I have is my judgments, there could be no discrepancy between what I judge ought to be a reason to act, and that which I judge as a reason to act. On this view there could be no gap which would allow α and β to be true at the same time:

(α) I acknowledge that I ought to take A to be the better reason.

and

(β) I take B to be the better reason.

But I have been arguing that the thought of this impossibility is the consequence of an overly limited view of what cognition involves.⁵⁰ Of course, one might ask why should we still attribute to the akratic agent the belief that he ought to act according to his reflective judgment, even though he doesn't, or, for this matter, the belief expressed by his reflective judgment. The answer to this question is already contained in the formulation of the problem of *akrasia*. The reasons to attribute to the agent such a belief are the same that lead us to think that the agent will act against her own best judgment: her sincere assent to this proposition, and perhaps, more important, the fact that attributing this belief to the agent will make best sense of her actions. After all, the agent's assent to a certain judgment from the reflective perspective will let us understand not only a single action, but much of her behavior, including her regrets towards and her attempts to overcome instances of *akrasia* in her behavior. It is indeed puzzling how it is possible that an agent will not follow her judgment from the reflective perspective, and it is this possibility that I hope we can now better understand.

One might also object that the akratic agent is not, under my description, strictly speaking irrational. For, after all, she does act according to the

⁵⁰ Charlton gives the clearest expression to the view that this gap, although possible in theoretical reason, is not possible in practical reason. Charlton claims that this difference is accounted for by the fact that "judging and not judging theoretically is not within our power. Practical judgment, on the other hand, is inseparable from action. That means that it is, in a way, in our power" (*Weakness of Will*, p. 126). My account of the distinction between direct and oblique cognition, and of the nature of reflective cognitions aims exactly at explaining why I do not make the judgment that was within my power to make. I also disagree with Charlton that the theoretical counterpart is easier to account for. As I argued earlier, it is important to distinguish the claim that judgment is a free activity from the claim that we can judge however we want. Indeed, we would find awkward (if not incomprehensible) if someone said to us: "I agree that if p then q , and, of course, I know how to use *modus ponens*, so I know I must conclude q . However I find myself not concluding that p , and what I theoretically judge is not within my power." We also need an account of how and when it is possible to fail to judge theoretically as one believes one ought. The virtue of the Cartesian approach to *akrasia* is that it can explain both.

reasons by which she has been persuaded. Moreover, this comes close to a denial of the phenomenon. The akratic agent, under my account, does not act against her reasons; she follows the reasons by which she is persuaded. However, it is not hard to locate the irrationality of the akratic agent. For she recognizes that she ought to judge that A is better than B, but she does not, and she also recognizes that it is within her power to judge as she ought to. This is in fact the archetype of irrationality: when an agent is fully responsible for things not being as he himself recognizes they ought to be from a reflective standpoint (or, all things considered). For the same reason, this account preserves the aspect of *akrasia* in question: the fact that the agent does not act as she recognizes that she ought to. Since she recognizes that she ought to judge that A is better than B, and that she is free to do A, she recognizes that she does not act as she ought and is able to act.

The following dialogue in Jane Austen's *Emma* is instructive:

Emma: "He may have as strong a sense of what would be right as you can have, without being so equal under particular circumstances to act upon it."

Mr. Knightley: "Then it would not be so strong a sense. If it failed to produce equal exertion, it could not be an equal conviction."⁵¹

The possibility that *Emma* raises might be thought to be a direct consequence of accepting the phenomenon of *akrasia*. But Mr. Knightley's correction is certainly appropriate. In *Emma*'s account it becomes hard to locate the responsibility of the agent, just as driving a wedge between motivation and evaluation leaves us with no distinction between *akrasia* and compulsion. Mr. Knightley's denial of an equal sense of duty should not be seen as a denial of the phenomenon, but rather as an insistence that even if the weak-willed person might know what is right, he cannot have "so strong a sense" of the right. Full understanding of one's duty leaves the free agent with no other option but to act, but this does not rule out the possibility of an understanding that has a lesser grasp of this same content, a grasp to which the grounds of the claim are not evident. And in this way we can save the phenomenon and yet pledge allegiance to the good old formula of the schools.

V

I would like to end with two observations that might clear up some confusion. First, I have been pressing the analogy between practical reason and theoretical reason, and claiming that we can understand *akrasia* by looking at the same phenomenon in its often overlooked theoretical incarnation. But it is important to see the limits of the analogy. In particular, it seems that I

⁵¹ *Emma*; New York: Penguin Classics, 1964; p. 119.

have a duty to explain why ‘theoretical *akrasia*’ is not as widespread as its practical counterpart.

Representation ought to be coherent in ways that evaluations need not be. A reason to believe *p* is generally a good reason to dismiss any proposition that is incompatible with *p*. This is not true of evaluating. Judging that I ought to pursue something else does not do away with the value of the action that I do not pursue.⁵² Incompatible values, values that cannot be simultaneously pursued, can coexist in way that incompatible worlds cannot. If I find out that pursuing an academic career is incompatible with becoming a millionaire, I must give up the *pursuit* of one these aims, but I need not thereby change my judgments about their value.⁵³ In the case of theoretical reason, if I find out that my beliefs that ‘Hesperus is a planet’ and ‘Phosphorus is not a planet’ are incompatible, I must give up holding one of them. A more important, and certainly not unrelated, fact is that in the case of theoretical reason, an adequate understanding of why a certain proposition is true also amounts to an understanding of why the claims that are incompatible with this proposition are false. But in the realm of practical reason, when one rules out a source of value as incompatible with what one takes as a more important project, this realization often does not do much to extinguish the appeal of the alternative ways of presenting the good. Practical wisdom is thus acquired through a more complicated process of education. Once one tries to structure one’s sensibilities so that they constitute a coherent and plausible view of the good, one has taken only one step in the process of silencing the opposing views. Having understood that we should not eat fatty foods, we often find the need to make this knowledge clear by reading about the gory details of how they contribute to the clogging of arteries, and so on. In the same way, advancing a political cause involves not only arguing, but, for example, making the presence of oppression and suffering vivid; this can produce not only the conviction that is often already there, but also the vivid understanding that competing considerations cannot carry much weight.

Secondly, any account of *akrasia* has to walk a thin line, which might be the reason why *akrasia* is such a puzzling phenomenon. In trying to account for *akrasia* we must avoid making the agent *wholly* incomprehensible. In this way, I argued that Davidson’s account of *akrasia* was unsatisfactory, because it either attributed to the agent an incredible form of stupidity or robbed her of

⁵² Wiggins makes a similar point in “Weakness of Will, Commensurability and the Objects of Deliberation.”

⁵³ These two evaluations are merely, we may say, *extrinsically* incompatible. There could be a world in which one could become a millionaire by pursuing a philosophical career. It is a contingent fact about our world that these two values turn out to be incompatible. But it might be the case that even evaluations that are *intrinsically* incompatible—evaluations whose objects are incompatible necessarily—can be held at the same time. Ideals such as friendship and impartiality might lead to evaluations that are incompatible in this way.

her freedom. But an account of *akrasia* cannot go as far as to make the agent *completely* intelligible, at least not in the sense of making the agent reasonable. As a form of irrationality, *akrasia* cannot be made reasonable.⁵⁴ What a philosophical account of *akrasia* must do, and what I purport to have done, is to explain what the akratic agent's temptation is, but not why he succumbs to the temptation instead of resisting it. To explain the possibility of *akrasia* is to show how a certain form of irrational behavior is possible and even tempting, but, of course, it cannot be to explain how a form of irrational behavior is the inescapable consequence of reasonable deliberation. To distrust our reflective judgment because it is too abstract or complex is irrational. But the obliqueness of certain reflective judgments let us see how we might fail to be persuaded by what we recognize ought to persuade us.

It is important to bear this point in mind to see whether an objection that David Pugmire raises against Wiggins's account of *akrasia* could also be raised against mine. Pugmire suggests that Wiggins's use of a notion of incommensurable forms of value to clarify the question of weakness of the will fails to make the phenomenon less perplexing. According to Pugmire, the fact that the lesser good holds a special charm could not help explaining akratic behavior:

The old problem remains: *How* can it happen? How could the sway of a special charm prevail over that of the *akrates*'s perceived larger interests *once* the two have been clearly appreciated against one another?⁵⁵

If my account of weakness of the will is right, the perplexity lies in an ambiguity in the last clause of Pugmire's remark. If by 'clearly appreciated', it is meant that the agent has an *unimprovable* grasp of the better course of action (or way of living), the description is false of the akratic agent. If his reflective judgment is grounded only on direct cognitions or that he has 'silenced' the claims of the competing perspectives, then there is indeed no room for *akrasia*. As Mr. Knightley points out, such a perfect sense of the right would ensue in action (unless step-motherly nature interferes). But if all that is meant is that the agent is convinced that on reflection the alternative course of action is better, the incommensurability of values leaves room for the possibility that this awareness is oblique, which in its turn, leaves room for the possibility that such an awareness could fail to persuade the agent. Seeing that the superiority of the alternative is clearly appreciated in this

⁵⁴ Davidson also points out this feature of a philosophical account of *akrasia*: "if we explain it [irrationality] too well, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diagnose irrationality by withdrawing the background of rationality needed to justify any diagnosis at all" ("Paradoxes of Irrationality", p. 303. See also "How is Weakness of the Will Possible", p. 42.)

⁵⁵ "Perverse Preference: Self-Beguilement or Self-division?" *Canadian Journal of Philosophy* 24 (1994), p. 93.

latter sense precludes us from seeing the akratic behavior as an instance of sensible behavior. But seeing that the superiority of the alternative is not clearly appreciated in the former sense, or even that it could not be clearly appreciated in the former sense, lets us see where the temptation lies.⁵⁶

⁵⁶ I am indebted to Annette Baier, Robert Brandom, Stephen Engstrom, Hans Lottenbach, John McDowell, Jennifer Nagel, Fred Schueler, Amy Schmitter, an anonymous referee for this journal, and audiences at the University of Pittsburgh and at Northwestern University for helpful comments on earlier versions of this paper.