3. There is a version of internalism (Tolhurst, 1998) that argues that what leads to motivation is the quasi-perceptual recognition that something "seems" wrong or right. If moral judgment requires this sort of nondeliberational perception of states of affairs as right or wrong, and if it can be shown that such perceiving is intrinsically connected to conative reactions, then perhaps a version of internalism would prevail in the face of VM evidence. Whether that would do the work traditionally supposed by internalists remains to be seen because it would link appearance rather than judgment to motivation.

4. Note that Kennett and Fine claim that "the central notion of a normative requirement [is] one that persists in the absence of inclination" (this volume, p. 178). If so, then grasping this normative requirement, or judging on the basis of it, ought to be possible in the absence of inclination. This is simply the externalist's claim.

4.2 The Truth about Internalism

Michael Smith

In her 2003 paper on ethical judgments and acquired sociopathy, Adina Roskies argues against the following very strong version of motive internalism (strict motivational internalism, SMI):

It is conceptually necessary that if an agent judges that she morally ought to ϕ in circumstances C, then she is motivated to ϕ in C.

According to Roskies, SMI is implausible, given what we know about patients with ventromedial frontal lobe damage. These patients appear to make moral judgments with full understanding since their use of moral language is just as complex and competent after their VM damage as it was before. However, after their VM damage, and in contrast to the way they were before the damage, they lack moral motivation. It is thus not just logically possible for an agent to judge that he or she morally ought to act in a certain way and not be motivated, this is what we have come to expect when the agent in question has suffered VM damage. So herein lies the truth about internalism: If SMI is the only version of internalism worth discussing, then we have good empirical reasons to believe that internalism is false.

Jeanette Kennett and Cordelia Fine take Roskies to task in their chapter. Their complaints are multiple, but for present purposes I will focus on just one: Roskies' assumption that SMI is the only version of internalism worth discussing. SMI says that it is literally impossible for an agent to make a moral judgment and yet not be motivated, that any failure of motivation is indicative of a failure of understanding. Although this strong claim is accepted by some—Kennett and Fine cite John McDowell—they insist that it is far more common for internalists to make much weaker claims about the connection between moral judgment and motivation. They therefore spend some time formulating their own preferred weaker version of internalism, a version that they think is more worthy of critical attention by opponents and which Roskies' argument leaves intact.

Kennett and Fine are, I think, right that SMI posits a connection between moral judgment and motivation too strong to be credible. For one thing, SMI makes it hard to see how weakness of will—motivation contrary to better judgement—is so much as possible. Instead it seems to commit us to an implausible Socratic view of weakness of will as a defect, not of the will, but of the understanding. Like Kennett and Fine, I am, however, nonetheless attracted to internalism, so I applaud their attempt to formulate a weaker version, a version that is both immune to Roskies' criticisms and on which opponents of internalism might more profitably focus. However, the alternative version of internalism that they come up with is not the one that I would have proposed myself, so in this commentary I want to explain why and offer my own alternative (see also M. Smith, 1994, 1997). As it happens, the alternative I favor is one that Roskies herself considers briefly but dismisses in her 2003 paper. At the end I explain why I think that Roskies' rejection of this weaker version of internalism was too hasty.

The version of internalism that Kennett and Fine propose is weaker than SMI in two ways. First, it replaces the conceptually necessary connection posited by SMI with a ceteris paribus connection. Second, it restricts the circumstances in which this weaker ceteris paribus connection between moral judgment and motivation is supposed to apply to the circumstances in which the judgment itself pertains. Their preferred version of motive internalism can thus be stated as follows (K&FMI):

Other things being equal, if an agent makes the in situ judgment that she ought to ϕ in circumstances C—that is, if she judges that she ought to ϕ in circumstances C, believing herself to be in those circumstances—then she is motivated to ϕ.

However, I see two main problems with K&FMI. The first concerns the restriction to in situ judgments. The second concerns the weakening of the connection to one that holds only ceteris paribus.

Kennett and Fine provide the following argument for the restriction to in situ judgments.

> I can surely believe that I ought to keep my promise but fail to form the in situ judgment that I ought to keep my promise. Maybe my belief or my promise isn't foregrounded in my deliberations about what to do. Maybe I fail to notice that what I'm planning to do—go to the football game this afternoon, say—would be inconsistent with keeping the promise I made two weeks ago to meet you in the mall at 3:00 on Saturday the 22nd. If I forget my promise to you, or I don't notice that the time to keep it is now, then, although I do believe I ought to keep my promises, including this one, I fail to form a judgment about what I ought to do right now, and so I fail to meet you. Of course I ought to form the judgment, but my failure to do so is not a failure of motivation. As we have described it, it can be a failure of attention or memory or inference. We take it that this kind of mismatch between moral belief and motivation to act wouldn't be enough to refute motive internalism. The relevant judgments are first person, in situ. (this volume, p. 182)

However, the conclusion of Kennett and Fine's argument, which is that we should restrict internalism to in situ judgments, does not follow from the premises they provide.

As stated, SMI requires that the content of people's motivations match the content of their moral judgments. When I believe that I ought to keep my promise to you now, SMI requires that I be motivated to keep my promise to you now. When I believe that I ought to keep my promise to meet you in the mall at 3:00 pm on Saturday the 22nd, SMI requires that I be motivated to meet you in the mall at 3:00 pm on Saturday the 22nd, and so on. Kennett and Fine point out, perfectly correctly, that if I believe that I ought to keep my promise to meet you in the mall at 3:00 pm on Saturday the 22nd but I don't believe that it is now that day and time, then I may not be motivated to meet you in the mall now. They also point out, again perfectly correctly, that this combination of moral belief and failure of motivation—believing that I ought to keep my promise to meet you in the mall at 3:00 pm on Saturday the 22nd, but not being motivated to meet you in the mall now—is not a counterexample to internalism. However, since it isn't a case in which there is a mismatch between the content of my moral belief and my motivation, it isn't a counterexample to SMI either. The argument they give thus provides us with no reason at all to think that SMI should be understood as making a claim about in situ judgements only.

What would motivate restricting SMI to in situ judgments? To motivate that restriction, Kennett and Fine would need to explain why we should suppose, on the one hand, that when I believe that I ought to meet you in the mall at 3:00 pm on Saturday the 22nd, I need not be motivated to meet you in the mall at 3:00 pm on Saturday the 22nd, even though when I believe that I ought to meet you in the mall now, I must be motivated to meet you in the mall now. It is tempting to think that no such explanation could be provided. The in situ moral belief would, after all, appear to be derived by putting the belief that I ought to meet you in the mall at 3:00 pm on Saturday the 22nd together with the belief that it is now 3:00 pm on Saturday the 22nd. However, if the in situ belief is derived from a non-in situ belief and a belief about what day and time it is now,

then it is surely plausible to suppose that what explains the necessary connection between the in situ belief and the in situ motivation is the perfectly general claim that moral judgments and motivations must have matching contents, and that the in situ motivation is therefore also derived by putting together the non-in situ motivation—the motivation to meet you in the mall at 3:00 pm on Saturday the 22nd—together with that same belief about what day and time it is now. So not only do Kennett and Fine fail to provide a convincing argument for the restriction of internalism to in situ judgments, the restriction itself looks very difficult to motivate.

The second problem with K&FMI concerns the weakening of SMI so that the conceptually necessary connection is replaced by a ceteris paribus connection. The problem with this particular way of weakening SMI emerges if we look more closely at SMI itself. SMI is false if the connection between moral judgment and motivation is contingent, even if, as a matter of fact, the connection is nomically necessary. Suppose, for example, that there is a contingent psychological law connecting the psychological state that underlies moral judgment with motivation. In that case the connection between moral judgment and motivation would be contingent but, as a matter of fact, nomically necessary. SMI would be false even though, as it happens, we never find someone making a moral judgment without being correspondingly motivated.

Once we notice the possibility of such a contingent yet nomically necessary connection between moral judgment and motivation, the crucial question to ask is whether such a connection would vindicate the truth of some version or other of internalism. The answer, I take it, is that it would not. This is because the mark of internalism, whether the internalism in question is of the strong kind posited by SMI or of some weaker kind, must surely be that it posits some sort of conceptually necessary connection between moral judgment and motivation. Internalism is, after all, supposed to function as an a priori constraint on what is to count as a moral judgment. The connection between moral judgment and motivation must therefore hold in virtue of the content of the moral judgment itself. It cannot be a connection that we discover empirically by uncovering a contingent psychological law.

The problem with K&FMI should now be apparent. K&FMI posits a connection between moral judgment and motivation that holds other things being equal. What is it for other things to be equal? Suppose that there were a contingent psychological law connecting the state that underpins moral judgment with motivation. In that case, other things would be equal when the state that underpins moral judgment did its causal work. So if

there were a contingent psychological law connecting the state that underpins moral judgment with motivation, then when other things are equal, someone who judges that they ought to act in a certain way would be motivated to act in that way. The existence of such a contingent psychological law would thus be sufficient to guarantee the truth of K&FMI. However, it would be insufficient to guarantee the truth of internalism, understood as a thesis that holds of conceptual necessity. So K&FMI doesn't state a version of internalism at all.

Of course, this leaves us with a problem. For if SMI states a version of internalism that is too strong to be credible, and if K&FMI doesn't state a version of internalism at all, then how exactly are we to formulate the weaker version of internalism on which opponents should focus? My own view is that the following weaker thesis (WMI) captures what's crucial:

It is conceptually necessary that if an agent judges that she morally ought to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational.

WMI says that what is supposed to be a conceptual truth is not that agents are motivated to do what they judge themselves morally obliged to do—this is the claim SMI makes—but rather that a failure to be so motivated is a form of practical irrationality. Nor should it be surprising that SMI should need to be weakened in this way, for the earlier criticism of SMI was that it didn't allow for the possibility of weakness of will as a genuine defect of the will. The difference between SMI and WMI is precisely that it allows for this possibility. Weakness of the will is, after all, just the name we give to a kind of practical irrationality that explains why someone judges that they morally ought to act in a certain way without being motivated to act that way.

In her 2003 paper, Roskies in effect considers this formulation of internalism, but decides that it is not worth discussing. Her complaint is that without a substantive characterization of what it is to be practically rational, WMI is trivially true. Suppose, for example, that a defender of WMI refuses to provide such a substantive account and instead simply stipulates that an agent is practically irrational whenever she judges that she morally ought to ϕ in C but isn't motivated to ϕ in C. (It might be thought that I came close to doing that just now when I characterized weakness of the will.) Roskies' objection is that no one could object to WMI, so understood; it is trivially true, given the stipulation. So for WMI to be worth discussing, a defender of WMI must therefore provide a substantive account of what

it is to be practically rational. Since no such account has been provided, Roskies concludes that WMI is best ignored.

I am not quite sure what Roskies thinks the defender of WMI needs to provide by way of a substantive account of practical rationality. On certain understandings, however, it seems to me that we should be skeptical about the truth of WMI given any such substantive account. Suppose, for example, that we substitute the kind of operational definition of what it is to be practically rational that a medical practitioner or a social worker might use in figuring out whether someone is capable of making autonomous choices: "able to describe the alternatives she faces, talk sensibly about their relative merits, and make a choice without getting flustered or overemotional." The trouble with WMI, given this operational definition of what it is to be practically rational, is obvious. Many people who aren't motivated to do what they judge they morally ought to do are nonetheless able to describe their alternatives, talk sensibly about their relative merits, and choose without getting flustered or overemotional. So, if we understand WMI in terms of such an operational definition of what it is to be practically rational, then WMI is no more credible than SMI.

It isn't clear why defenders of WMI should accept that their alternatives are either to stipulate a meaning for being practically rational (in which case WMI is trivially true) or to provide a substantive characterization of (say) the operational kind just mentioned (in which case WMI is obviously false). In other domains there is plainly a third kind of alternative. Consider, for example, the following claim about theoretical rationality (MORTAL):

If someone believes that Socrates is a man and she believes that all men are mortal, then either she believes that Socrates is mortal or she is theoretically irrational.

Again, under any plausible operational definition of what it is to be theoretically rational, MORTAL looks bound to turn out false. Suppose, for example, that we count people as theoretically rational if and only if they score higher than the average on the Scholastic Aptitude Tests (SATs; these are the standard assessment tests used to determine relative standing among high school students who compete for college entry in the United States). The trouble with MORTAL given this understanding of what it is to be theoretically rational is plain, for someone could easily score higher than the average on the SATs and yet fail to believe that Socrates is mortal when she believes that Socrates is a man and believes that all men are mortal. MORTAL, so understood, is thus plainly false. Moreover, any similar operational definition of what it is to be theoretically rational looks like it would make MORTAL turn out similarly false, for the simple reason that the respect in which someone who fails to believe that Socrates is mortal when she believes that Socrates is a man and that all men are mortal is irrational is precisely this very respect. This particular combination of belief and lack of belief—believing that Socrates is a man, believing that all men are mortal, but not believing that Socrates is mortal—constitutes an instance of theoretical irrationality. An operational definition, by contrast, at best identifies some feature that roughly correlates with such instances of theoretical irrationality.

Does this mean that the defender of MORTAL is reduced to stipulating what is to count as theoretically rational? That does not seem to be an accurate description of what is going on either. In order to see why, contrast the situation of someone trying to defend MORTAL with someone trying to defend the following claim (CAPITAL):

If someone believes that Canberra is the capital of Australia, then either she believes that Vienna is the capital of Austria or she is theoretically irrational.

If she is to succeed, then it seems that the defender of CAPITAL has no alternative but to stipulate that, as she uses the term "theoretically irrational," someone will count as theoretically irrational when she believes that Canberra is the capital of Australia but doesn't believe that Vienna is the capital of Austria. She has no alternative because she can provide no account of why this particular combination of belief and lack of belief constitutes an instance of theoretical irrationality in any ordinary sense. The defender of MORTAL, by contrast, can provide such an account. The explanation, very roughly, is that being theoretically irrational in the ordinary sense is a matter of a failure of sensitivity in the formation of your beliefs to what you take to be reasons for belief. The combination of belief and lack of belief that the defender of MORTAL thinks constitutes an instance of theoretical irrationality is an instance of just such an insensitivity, whereas the combination of belief and lack of belief that the defender of CAPITAL thinks is an instance of theoretical irrationality is not. That Socrates is a man and that all men are mortal are, by the lights of someone who believes these things, reasons for believing that Socrates is a man, as there is, by the lights of the person who has these beliefs, an inferential connection. But that Canberra is the capital of Australia is not,

by the lights of someone who believes this to be so, a reason for believing that Vienna is the capital of Austria. There is not, by the lights of the person who has this belief, any such inferential connection.

The question for the defender of WMI is whether she can make a similar move. Can she explain why this particular combination of belief and lack of motivation—someone's believing that she morally ought to ϕ in C and yet lacking any motivation to ϕ in C—constitutes an instance of practical irrationality in an ordinary sense? If she can, then it isn't appropriate to describe her as merely stipulating what she means by "being practically irrational." My suggestion is that the defender of WMI can provide such an explanation. The explanation comes in three stages. Moreover, and importantly, at each stage the explanation remains faithful to the observation made earlier that according to internalists the connection between moral judgment and motivation must hold in virtue of the content of the moral judgment itself.

At the first stage the defender of WMI must argue that what it is that someone believes when she believes that an agent morally ought to ϕ in C is that ϕ-ing in C is that action in C, among the agent's alternatives, that uniquely maximizes value. At the second stage she must argue that what it is that someone believes when she believes that p has value is that p is something that she would want if she had a maximally informed and coherent and unified desire set. And at the third stage she must argue that the following combination of belief and lack of desire—someone's believing that she would want that p, if she had a maximally informed and coherent and unified desire set, but lacking any desire that p—constitutes an instance of practical irrationality in an ordinary sense. Although I fully admit that the claims made at all three stages of this explanation are controversial, it is, I think, important to note that the defender of WMI is in fact providing arguments at every stage and hence is not merely stipulating.

Consider, for example, what the defender of WMI might say in defense of the most controversial claim of all, the claim she makes at the third stage. What is especially striking about this claim is that it too trades on the idea that an agent's irrationality is a matter of her insensitivity to reasons. In this case, though, the reason the agent has, at least by her own lights, is a reason to desire that p: by her own lights, that she would want that p if she had a maximally informed and coherent and unified desire set provides her with a reason for wanting that p here and now, as she is actually. To be sure, the force of the putative reason is derivative. The putative reason-providing force of the complex fact derives entirely from the features of the agent's individual desires and their contents and the relationships among them and information that is supposed to make it the case that she would indeed want that p if she had a maximally informed and coherent and unified desire set. It provides the agent with a reason nonetheless (although contrast Scanlon, 1998, chap. 2).

Indeed, the practical irrationality in this case—the failure to be sensitive to the reasons for wanting—seems to be on the same level as the *theoretical* irrationality—the failure to be sensitive to one's reasons for believing—manifested by someone who believes that she would believe that p if she had a maximally (otherwise) informed and coherent and unified belief set, but who fails to believe that p. By her own lights, after all, the fact that she would believe that p if she had a maximally (otherwise) informed and coherent and unified belief set provides such an agent with a reason to believe that p. Again, the force of this putative reason is derivative. The putative reason-providing force of the complex fact derives entirely from the features of the agent's beliefs and their contents and the relationships among them that are supposed to make it the case that the agent would indeed believe that p if she had a maximally (otherwise) informed and coherent and unified belief set. It provides the agent with a reason nonetheless. The only difference between the two cases is that in the second one the agent manifests a failure to believe in accordance with what, by her own lights, are reasons for believing, whereas in the former case the reason is a reason for wanting.

The upshot is thus that Roskies was far too quick in her rejection of WMI. WMI states a weaker connection between moral judgment and motivation than that implied by SMI. Moreover, as we have seen, this weaker connection can be supported by arguments, not by a mere stipulation. Both defenders and opponents of internalism would therefore do best to confront these and similar arguments for WMI head-on; therein lies the truth about internalism.