Knowledge of Moral Incapacity

Ryan Cox

ABSTRACT

Are the limits on what we can do, morally speaking—our "moral incapacities" as Bernard Williams calls them—imposed on us from within, by reason itself, or from without, by something other than reason? Do they perhaps have their source in the will, as opposed to reason? In this essay, I argue for a theory of moral incapacity on which our moral incapacities have their source in reason itself. The theory is defended on the grounds that it provides the best explanation of our knowledge of our own moral incapacities. I argue that just as an agent's reflective commitments play an essential role in the explanation of their knowledge of their moral incapacities, they play an essential role in the explanation of moral incapacities themselves. Since reflective commitments are rational commitments, and rational commitments have their source in reason, moral incapacities have their source in reason itself. The theory of knowledge of moral incapacity offered in this essay draws on elements of Richard Moran's "deliberative" theory of self-knowledge and elements of that theory are used to offer a theory of moral incapacities which extends and improves on Bernard Williams' "deliberative" theory of moral incapacities. The resulting theory is then defended against objections and alternatives.

1 Introduction

My topic in this essay is what Bernard Williams calls "moral incapacity" (Williams 1995). I propose to approach this topic by means of a closely related topic, namely, *knowledge* of moral incapacity. I will first introduce each topic, and then explain the connection between them.

The topic of *moral incapacity* itself can be introduced by examples. There are familiar examples from history, like Luther's "I can do no other, here I stand."[1] There are examples from literature, like the example of Huck Finn from Twain's *Huckleberry Finn* who cannot turn his friend Jim in to the authorities (Taylor 2001, 59). And, of course, there are examples from philosophy. Williams gives the example of a man who cannot gossip to reporters about his wife's problems because he would find it "disloyal, shabby, etc" (Williams 1995, 50) and Harry Frankfurt gives an example of a mother who cannot give up her child for adoption (Frankfurt 1999, 111). On the face of it, what characterises all of these cases is that an agent cannot do something, in a perfectly ordinary sense of "cannot", and their incapacity bears some relation to the agents' moral character (Williams 1995, 46).

The related, but less familiar, topic of *knowledge of moral incapacity* can be introduced by reflecting on a remarkable feature of our knowledge of our moral incapacities, namely, that we seem to have a distinctive way of knowing about our own moral incapacities: the way we have of knowing about our own moral incapacities is distinct both from the way we have of knowing other kinds of incapacities of our own and of the way we have of knowing the incapacities, moral or otherwise, of others. Williams has drawn attention to the fact that first-personal conclusions of the form "I cannot ϕ", when they express moral incapacities, may be deliberative conclusions, based not on evidence that one cannot ϕ, but on practical considerations (Williams 1995, 47). The point here is that one can reach conclusions about one's own moral incapacities not by reflecting on psychological evidence bearing on what one can or cannot do, but by reflecting on reasons for and against performing the very action in question. The contrast is with how one reaches conclusions about other incapacities of one's own and how one reaches conclusions about the incapacities, whether moral or otherwise, of others. On the face of it, then, our knowledge of our own moral incapacities appears to be much like our knowledge of our own attitudes, in the general sense that we seem to have a distinctive way of knowing about them and

---

[1] The example is discussed in Williams (1995) p. 48; Taylor (1995) p. 276; Frankfurt (1998b) p. 86; Van den Beld (1997) p. 529

in the specific sense that we seem to be able to reach conclusions about our own moral incapacities and attitudes not by reflecting on things which are evidence that we have the incapacities and attitudes in question but by reflecting on reasons for and against performing the action and having such attitudes. While this feature of Williams's account has been noted by commentators (Van den Beld 1997, 527–28; Gaita 2004, 107; Watson 2002, 142), it has not received significant attention, and the connection between this form of self-knowledge and other forms of self-knowledge has not been drawn.

These two topics—*moral incapacity*, and *knowledge of moral incapacity*—are related, since, if these observations about knowledge of moral incapacity are correct, they put significant constraints on a theory of moral incapacity. Moral incapacities must be such than an agent can know their own moral incapacities by such a distinctive means. It is a mark against a theory of moral incapacity if it cannot explain this, and it is a mark in favour of a theory if it can. [2] Moreover, going in the other direction, if these observations are correct, a general theory of self-knowledge must be able to account for knowledge of moral incapacities. It is a mark against a general theory of self-knowledge if it cannot account for knowledge of moral incapacities, and it is a mark in favour of a theory if it can.[3] While my focus in this essay will be on the first issue, much of what I say here is directly relevant to general theories of self-knowledge and I will draw attention to these connections throughout the essay without pursuing them in detail.[4] My aim, in this essay, then, will be to leverage observations about *knowledge* of moral incapacity in arguing for a particular theory of moral incapacity. More precisely, I will use elements of Richard Moran's "deliberative" theory of self-knowledge to modify and extend elements of Williams's "deliberative" theory of moral incapacity and to offer an additional argument for it.[5] I

---

[2] In my view, it is a mark against Frankfurt's theory of moral incapacity that it cannot explain knowledge of moral incapacities. To make the case for this claim would require a careful explication of Frankfurt's position and an examination of alternative explanations of knowledge of moral incapacity which I am not able to undertake here. In short, the difficulty for Frankfurt's position stems from Frankfurt's insistence that moral incapacities are a volitional matter and that the volitional states involved in moral incapacities in no way depend on an agent's reasons. (See Frankfurt (2002) p. 219). This effectively rules out an appeal to the kind of transparency based theories of self-knowledge considered in this essay. See (Frankfurt 1998a).

[3] There is debate over how general a theory of self-knowledge must be. See (Finkelstein (2003) p. 162; Bar-On (2004) p. 144; Boyle (2009) pp. 140–141; Byrne (2011b) pp. 212–214]. That debate is mainly concerned with whether a theory of self-knowledge must be general with respect to knowledge of sensory—perceptual states and other phenomenally conscious states—and non-sensory states—attitudes. I am not claiming that a theory of self-knowledge must be fully general. But it would be a mark in favour of a theory if it could be general with respect to knowledge of our attitudes and knowledge of moral incapacity. And I focus on this case here. There is a growing literature concerning just how far transparency accounts of self-knowledge can be applied. See, for instance, (Ashwell 2013; Byrne 2011a, 2011b; Setiya 2011; Kloosterboer 2015; Cox 2018; Keeling 2019b, 2019a).

[4] As the discussion to follow will make clear, so-called transparency theories of self-knowledge are best placed to account for knowledge of moral incapacity. I take this to be a mark in their favour. For transparency theories, see Evans (1982), Moran (2001), Fernandez (2013), and Byrne (2018).

[5] I focus on Moran's theory since, as I will argue, it has a feature that other transparency theories of self-

argue that Williams does not do enough in his discussion to explain how it is that we can come to know our moral incapacities by a distinctive means and so does not do enough to defend his theory of moral incapacities. Using the notions of "deliberative commitment" and "first-person authority" developed in Moran's theory, I offer an argument for a theory of moral incapacities which improves on Williams's theory and his arguments for it. Insofar as Moran's "deliberative" theory of self-knowledge is plausible as a general theory of self-knowledge, we have good reasons for developing a theory of knowledge of moral incapacity and a theory of moral incapacity which draw on this theory. At the same time, to the extent that the theory can explain the facts about knowledge of moral incapacity, this lends support to Moran's theory as a general theory of self-knowledge.[6]

2 Locating the Topic

Williams initially locates the topic of moral incapacity by claiming that it is "the kind of incapacity that is in question when we say of someone, usually in commendation of him, that he could not act or was not capable of acting in certain ways" and by claiming that moral incapacities, unlike other incapacities, are closely related to "what sort of person the agent is", that is, to their character (Williams 1995, 46). Luther's inability to do otherwise, Huck's inability to turn Jim in, the husband's inability to gossip to reporters about his wife, and the mother's inability to give up her child, all seem to have these features. They say something about the kind of agent each is, something about their moral character.

Williams then notes that in thinking about moral incapacities we can think about them in both their first-personal aspect and their third-personal aspect.[7] In thinking about them

knowledge lack and which is crucial for explaining knowledge of moral incapacity.

[6] There are many critics of Moran's "deliberative" transparency theory. Some are critical of transparency theories in general. See, for instance, (Finkelstein 2003; Bar-On 2004; Way 2007; Lawlor 2009; Gertler 2011; Paul 2014; Cassam 2015)) Others are critical of Moran's particular version of the transparency theory. See, for instance, (Byrne 2018). If these critics reject Moran's theory, then they must either reject the observations I make in the text about knowledge of moral incapacities or find an alternative and better explanation of knowledge of moral incapacities. I take it to be a significant mark in favour of Moran's particular version of the transparency theory that it can so straightforwardly explain knowledge of moral incapacities. Others have developed transparency theories which are largely sympathetic to Moran's theory but which modify it in certain ways. See (Boyle 2009, 2011a, 2011b, 2019; Hieronymi 2006; Hieronymi 2009). Insofar as these theories are structurally similar to Moran's view they may also be extended to explain knowledge of moral incapacities. Further work would need to be done to compare the relative advantages and disadvantages of these theories.

[7] Not all moral incapacities have both a first-personal and a third-personal aspect. As Williams notes, there are cases were someone is incapable of doing something simply because an action of the relevant kind is not part of their practical vocabulary. In such cases, we may be able to say in the third person that the agent is incapable of doing the thing in question. But that thought will obviously be unavailable to them since they are unable to think of the action at all. Following Williams, I focus here on incapacities which have both a first-per-

in their third-person aspect we note that claims about moral incapacity, claims of the form 'He cannot', are not merely claims of deontic modality, but are, like other claims about incapacity, claims of circumstantial modality, are claims about ability. As Williams notes, ascriptions of moral incapacities "are connected with prediction and understanding in some of the ways that physical, and other psychological, incapacities are. In particular, they sustain inferences of the form *esse ad posse* and, significantly, *non posse ad non esse*" (Williams 1995, 47).

It is in thinking about moral incapacities in their first-personal aspect that we notice something remarkable about them, namely, that first-personal ascriptions of moral incapacities may be the conclusion of practical deliberation. As Williams puts it:

> There is also a first-personal 'I can't', which is connected with the same kind of subject matter as third-person ascriptions of moral incapacity, and is related to deliberation. The 'I can't' may express the conclusion of a deliberation: my practical question has been whether to ϕ, and after rehearsing the moral and other reasons I come to the conclusion that I cannot ϕ. (Williams 1995, 46)

I will have more to say in a moment about what exactly Williams is saying here about how we arrive at first-person ascriptions of moral incapacity. For now, I just want to draw attention to how the first-person and third-person aspects of moral incapacities are related. We can think of Williams as making the following three *prima facie* plausible claims about ascriptions of moral incapacities:

1. Third-personal ascriptions of moral incapacities are ascriptions of genuine incapacities or inabilities: they involve a kind of circumstantial modality.

2. Third-personal ascriptions and first-personal ascriptions of moral incapacities are about the same subject matter: they involve the same sense of 'cannot'.

3. First-personal ascriptions of moral incapacities can be arrived at on the basis of practical deliberation.

While each of these claims are prima facie plausible, their combination amounts to something remarkable. If third-person ascriptions of moral incapacities are ascriptions of genuine incapacities or inabilities, and first-person ascriptions involve the same sense of 'cannot' and can be arrived at on the basis of practical deliberation, then we have a distinctive way of arriving at ascriptions of genuine incapacities or inabilities. Our

sonal and a third-personal aspect. See Williams (1995) pp. 46–47 and Williams (1981) pp. 128–129 for discussion.

knowledge of our own moral incapacities—as genuine incapacities—is more like our knowledge of our own attitudes and actions than it is of our knowledge of other incapacities and our knowledge of the attitudes and actions of others. If our knowledge of our own moral incapacities is distinctive in this way, then this raises important questions about what moral incapacities are.

We can think of Williams as attempting to offer a theory of moral incapacities which vindicates all three of these claims. Indeed, he writes in connection with (3) that "when we understand better what kind of incapacity is in question here, we shall see that this is what we should expect" (Williams 1995, 48). Yet one might think that this remarkable implication is a reason to reject at least one of the claims: we cannot know facts about genuine incapacities on the basis of practical deliberation. A sceptic might argue then, that we should accept (1) and (2) and reject (3): since first-person ascriptions are about genuine incapacities, they must, like ascriptions of other kinds of incapacities, be based on similar kinds of evidence. Or they might argue that we should accept (1) and (3) and reject (2): since first-person ascriptions are not based on ordinary evidence, they cannot be about the same subject matter as third-person ascriptions of moral incapacities. Indeed, some theorists have questioned whether first-person ascriptions involve the same sense of 'cannot' as third person ascriptions (Gaita 2004, 107; Van den Beld 1997, 526). Williams is aware of the temptation here. He writes:

> Here someone will reach for the weapon of distinguishing senses, and will speak of there being two or more senses of 'cannot', that which signifies whatever rejection is embodied in the agent's deliberation, and that which expresses what one 'literally' cannot do. (Williams 1981, 127)

Of course, to claim that there is a special sense of 'cannot' involved in first-personal ascriptions of moral incapacities conflicts with the fact that third-personal and first-personal ascriptions seem to involve the same sense of 'cannot'. So if the sceptic wants to accept (2) and (3), they must reject (1), and hold that third-person ascriptions, contrary to appearances, do not ascribe genuine incapacities, but ascribe something else.

I take it that what Williams says about (1) is enough to make it initially plausible, and I assume that (2) is also independently plausible. We are yet to see the motivation for (3)—the claim that first-personal ascriptions of moral incapacities can be arrived at on the basis of practical deliberation. In the next section, I clarify and motivate (3).

We have now located the subject matter of a theory of moral incapacities. Before turning to the issue of knowledge of moral incapacities, I want to briefly remark on something else that Williams says about the subject matter. Williams claims that one important way that moral incapacities differ from other incapacities is that moral incapacities, unlike the others are "related to what sort of person the agent is...to his character" (Williams 1995, 46), they are "expressive of, or grounded in, the agent's character or personal dispositions" (Williams 1995, 47). As we will see, they are, according to Williams, particularly related to a person's dispositions to deliberate in certain ways. Now, it may seem that, in light of such remarks, we should expect the notion of character to play a significant role in a theory of moral incapacities. Williams rejects this suggestion, however. He writes: "I do not intend the notion of 'character' itself to be explanatory in this discussion; it helps, if at all, in locating what has to be explained (Williams 1995, 47). As Williams writes:"'character' does not contribute much to the explanation of moral incapacity; perhaps, rather, if we can shed light on moral incapacity we can shed some light on the idea of 'character'."[8].

3 Knowledge of Moral Incapacity

Consider Williams' claim about first-personal ascriptions of moral incapacities being arrived at on the basis of practical deliberation again: "'I can't' may express the conclusion of a deliberation: my practical question has been whether to ϕ, and after rehearsing the moral and other reasons I come to the conclusion that I cannot ϕ (Williams 1995, 48). Now consider an example to illustrate it. I may be deliberating about what to do to improve my essay on moral incapacities, and when the option of including a student's idea on the topic in my essay unacknowledged comes up, I immediately conclude that I cannot do it. Now, it is important to distinguish two distinct but closely related versions of this example. In one version, I do not seriously consider including the student's idea as an option and I simply arrive at the conclusion in the course of deliberation that I cannot do it. In another version, I may initially be able to consider it as an option, but may not in the end be able to choose to do it: in the course of deliberation I may recognise that it has a certain feature which means that I cannot treat it as a serious option and may then conclude that I simply cannot do it.[9] Either way, my conclusion is a *deliberative* conclusion—a

---

[8] (Williams 1995, 47). Also: "Incapacities can not only set limits to character and provide conditions of it, but can also partly constitute it" (Williams 1981, 103)

[9] c.f. "What I recognise, when I conclude in deliberation that I cannot do a certain thing, is a certain incapacity of mine. I may be able to think of that course of action, but I cannot entertain it as a serious option. Or I can consider it as an option, but not in the end choose it or do it" (Williams (1981) p. 128).

conclusion arrived at on the basis of practical deliberation, deliberation which involves considering reasons for and against various options. It is *not* the conclusion of a more theoretical kind of deliberation about what I am capable of doing, a kind of deliberation which would seek evidence for what I am capable of. As Williams writes:

> The deliberative conclusion is based, as decisions are, on considerations of the good, the useful, the obligatory and so on, and not on psychological information about myself, which is presumably what I would need in order to have a basis for ascribing an incapacity to myself. (Williams 1995, 48)

In arriving at the conclusion that I cannot include my student's idea unacknowledged, I do not consider psychological information about myself, about what kind of person I am, about my motivations, and so on. Rather my conclusion is based on considerations of the good, the useful, and the obligatory, or, in other words, it is based on the consideration of reasons *for* and *against* doing the thing in question, or reasons *for* and *against* treating doing the thing in question as an option in deliberation. It may be based on my recognition that including the idea unacknowledged would be a shabby or deceitful thing to do, for instance.

We can discern in Williams's remarks, then, several positive theses about knowledge of moral incapacities and one negative one. The positive theses are these: (i) the conclusion that I cannot do something may be the conclusion of practical deliberation about whether to do the thing in question (ii) and when it is, the conclusion is based on the consideration of reasons for and against doing the thing in question or on reasons for and against treating doing the thing in question as an option in deliberation. The negative thesis is this: (iii) the conclusion is *not* based on psychological information about myself.

What is remarkable about Williams's claims about knowledge of moral incapacity, among other things, is the similarity they bear to remarks about other kinds of self-knowledge. Analogous claims have been made about knowledge of our actions and attitudes. Many theorists claim that our knowledge of our future actions is not based on psychological information about ourselves, but is, rather, based on considerations of reasons for and against performing the action in question (Hampshire and Hart 1958). And many claim that our knowledge of our attitudes is not based on psychological information about ourselves, but is, rather, based on reasons for and against having those attitudes. In the case of belief, for example, it has been claimed that our knowledge of our own beliefs is based not on psychological evidence about ourselves but evidence for the truth of the

propositions we believe (Hampshire 1975, 59; Evans 1982, 225; Moran 2001, 62). To put it in the contemporary parlance, questions about our own actions and attitudes are *transparent* to questions about the world (Edgley 1969; Moran 2001).

For our purposes the analogy with knowledge of action is the most useful. Such knowledge is thought to be both *immediate* and to exhibit a kind of *transparency*. As Moran writes:

> [A] person *may* have a purely predictive basis for knowing what he will do, but in the normal situation of free action it is on the basis of his decision that he knows what he is about to do. In deciding what to do, his gaze is directed "outward," on the considerations in favor of some course of action, on what he has most reason to do. Thus his stance towards the question, "What am I going to do now?" is transparent to a question about what he *is to* do, answered by the "outward-looking" consideration of what is good, desirable, or feasible to do. (Moran 2001, 105)

So, the question of what I am going to do might be said to be transparent to the question of what *I am to do*. And to answer the latter question I must consider reasons for and against acting. In this sense, the question of what I am going to do is, at least in part, *transparent* to a question about the world. Williams's observation about knowledge of moral incapacity is analogous: the question of what I *can* and *cannot* do, is, in some sense, and to some degree, transparent to a question about the world. As Williams writes in the context of conclusions about moral incapacity: "[t]he thought that leads to them ... is not for the most part thought about oneself, but thought about the world and one's circumstances" (Williams 1981, 130). So, to put Williams's observations in the contemporary parlance: knowledge of moral incapacity is both *immediate* and exhibits a kind of *transparency*—it is immediate in the sense that it is not based on evidence or observation, and it is transparent in the sense that answering questions about moral incapacities often involves answering a question about the world.

Now, these observations about immediate and transparent self-knowledge raise a challenge: how could it be possible to answer a question about one subject-matter—for example, our future actions or our attitudes—by considering reasons which bear on a question about another subject matter: the good, the useful, the obligatory, the true, the desirable, and so on? As we will see in a moment, this challenge is a version of what is known as the two-topics problem for transparency theories of self-knowledge. A theory of self-

knowledge would have to give an account of how this is possible. In a prescient remark, Williams raises exactly this issue: "[T]hough it needs to be understood in philosophy, [it] is not a paradox: it must be true, not only of practical reasoning but more generally, that one finds out about oneself by thinking about the world that exists independently of oneself" (Williams 1981, 130). While it is no paradox, the question remains just *how* it could be true that one "finds out about oneself" in this way.

My aim in the next section is to introduce a theory of self-knowledge which is able to explain these observations about transparency and self-knowledge. Since the theory is—arguably—independently motivated as an explanation of other kinds of self-knowledge, the fact that it can offer an explanation of knowledge or moral incapacity, lends further support to the claim that our knowledge of our own moral incapacities is special in the way that Williams claims that it is.

4 A Theory of Knowledge of Moral Incapacities

In the previous section I argued that knowledge of moral incapacities is like knowledge or our actions and attitudes in that it is both *immediate* and exhibits a kind of *transparency*. While many contemporary theories of self-knowledge take the immediacy of self-knowledge to be the central explanandum of a theory of self-knowledge, theories differ in the importance they assign to the transparency observation.[10] Some theories downplay the significance of the observation, while others make it central to the theory of self-knowledge. I will draw on a theory of self-knowledge which makes the transparency observation central to self-knowledge, namely, Moran's "deliberative" theory of self-knowledge. In this section I will briefly explicate Moran's theory, drawing attention to the elements of his theory which will be central to explaining knowledge of moral incapacities and which I will draw on in developing a theory of moral incapacities.

4.1 Moran on Transparent Self-Knowledge

A crucial element of Moran's explanation of the immediacy and transparency of self-knowledge is the distinction between theoretical and deliberative questions. The distinction can be brought out, albeit imperfectly, by opposing interrogative forms like "Do I believe that Jane runs?", "Do I intend to go to the party?", "Am I afraid of John?"—all of

---

[10] Ryle famously rejected immediate as an explanandum for a theory of self-knowledge. Recent theories which have rejected the explanandum, at least for knowledge of the attitudes, include (Carruthers 2011) and (Cassam 2015). See also (Lawlor 2009) for the case of desire.

which intuitively express *theoretical* questions—with interrogative forms like "Am I to believe that Jane runs?", "Am I to go to the party?", and "Am I to be afraid of John?"—all of which intuitively express corresponding *deliberative* questions.[11] This intuitive difference corresponds to a psychological difference in the way the questions are answered: a theoretical question is "answered [merely] by discovery of the fact of which one was ignorant" and is "a response to ignorance of some antecedent fact about oneself" (Moran 2001, 58) whereas, in contrast, a deliberative question is "answered by a *decision* or *commitment* of some sort" and is "not [merely] a response to ignorance of some antecedent fact about oneself" (Moran 2001, 58, my emphasis). One takes a *deliberative* stance towards the question of whether one believes, say, that Jane runs, when one lets one's answer to this question be fixed by one's answer to the corresponding deliberative question, that is, the question of whether one is to believe that Jane runs.

According to Moran, in the case of belief, "[what] transparency requires is the deferral of the theoretical question"What do I believe?" to the deliberative question "What am I to believe?" and, in turn, "answering a deliberative question is a matter of determining what is true" (Moran 2001, 63). To defer the theoretical question "What do I believe" to the deliberative question "What am I to believe?" is to take a deliberative stance towards the question. One then answers the question "What am I to believe?" by determining what is true. It is important to recognise that Moran's explanation of transparency has a two-part structure. First, the theoretical question is deferred to the deliberative question. And then the deliberative question is answered, in the case of belief, by considering factors bearing on the truth of the belief. For Moran, then, "transparency is grounded in the deferral of theoretical reflection on one's state of mind to *deliberative reflection* about it" (Moran 2004, 424, my emphasis). As I will argue below, the two-part structure of Moran's theory is essential to explaining knowledge of moral incapacities.[12]

But how can it be legitimate to answer a question about one subject matter (one's state of mind) as if it were a question about a wholly distinct subject matter (about the mind-

---

[11] Critics of Moran's view often characterise deliberative questions as normative questions like "Should I go to the party" or "Ought I to go to the party" (Finkelstein 2003; Cassam 2015). Moran is explicit that this is not his position (Moran 2001, 145). In terms of the psychological contrast about to be drawn in the text, a deliberative question is answered by a decision or commitment of some sort, whereas a normative question is answered by arriving merely at a judgement, belief or evaluation.

[12] Critics of Moran's theory often overlook the two-part structure of Moran's explanation and attribute to him the view that one simply answers the question "Do I believe that Jane runs?" by answering the question of whether Jane runs. Some say that on Moran's view one takes a deliberative stance to the question "Do I believe that Jane runs?" just when one answers it by answering the question of whether Jane runs. See, for instance, (Byrne 2005). But this is not so as the text makes clear. Some alternative transparency theories of self-knowledge replace the two-part structure of Moran's theory with a one-part structure. See, for instance, (Fernandez 2013; Byrne 2018).

independent world)? This is the two-topics problem for transparency theories of self-knowledge. Moran's answer is that it is legitimate because what one's attitude *is* is, partly, if not entirely, rationally determined by the conclusion of deliberative reflection. This is a modest sense in which we can be said to have a kind of agency or first-person authority with respect to our attitudes. Insofar as our attitudes are rationally determined by our answers to deliberative questions, we can be said to have a kind of agency or *first-person authority* over our attitudes. As Moran writes: "A person is credited with first-person authority when we take the question of what he *does* believe to be settled by his decision as to what he *is to* believe" (Moran 2001, 134). That is, a person is credited with first-person authority when the question of what attitude he has is settled by his decision or commitment arrived at by deliberative reflection. When someone is committed to a particular answer to a deliberative question, we can call this a *deliberative commitment*.[13] To the extent that our attitudes and actions are rationally determined by our deliberative commitments, we can be said to have first-person authority over our attitudes and actions.[14] For Moran, then, the notions of deliberative commitment and first-person authority provide a solution to the two-topics problem.

4.2 A Transparency Theory of Knowledge of Moral Incapacities

Moran's account of self-knowledge can be extended in a straightforward way to provide an account of knowledge of moral incapacities. We saw earlier that the theoretical question "Am I capable of ϕ-ing?" is transparent to questions about the good the useful and the obligatory. So, for instance, the man in Williams' example might answer the question "Am I capable of having a friendly and relaxed conversation with the lover of my spouse?" by considering whether it would be a shaming, dishonest, lax, or grotty thing to do (Williams 1995, 49). However, in order to mirror the two-part structure of Moran's theory we must find the deliberative question that the theoretical question "Am I capable of ϕ-ing? is deferred to, and which, in turn is answered by considering certain features of

---

[13] I use 'deliberative commitment' in a stipulative sense to refer to the state of being committed to an answer to a deliberative question. It is standard in the literature on commitment to think of commitments as a species of intention. See, for instance, (Calhoun 2009, 615). I suspect that in the case of action deliberative commitment in my sense and deliberative commitment in this sense coincide. Indeed, an intention can be thought of as an answer to a question about what to do. For a view of this kind, see (Hieronymi 2006, 2009). However, since there are deliberative questions about belief and emotion, I do not think that deliberative commitments are a species of intention. If anything, intentions are a species of deliberative commitment. See (McGeer 1996) and (McGeer 2008) for development of a theory of self-knowledge which involves yet another sense of commitment.

[14] None of this is to say that our agency or authority is perfect. It may well be that although we have reached a positive answer to a deliberative question, we lack the corresponding belief. Moreover, we may posses a particular belief while not being committed to a positive answer a corresponding deliberative question, or while being committed to a negative answer.

ϕ-ing itself. The deliberative question here cannot be whether to ϕ, for that would not distinguish questions about what I *can* do from questions about what I am doing or what I will do. Moreover, there's an important sense in which, at least in many central cases, being morally incapable of ϕ-ing means that one will *not* deliberate seriously about whether to ϕ. I want to suggest, then that the relevant deliberative question is the question of whether to treat ϕ-ing *as an option* in deliberation. One can address *this* question without deliberating about whether to ϕ, without entertaining it as a serious option.

So, it seems that in the case of moral incapacities, the question of whether one is capable of doing something is, at least in part, transparent to the question of whether one is to treat doing the thing in question as an option: am I to even treat this as an option? But how can it be possible to answer a question about one subject matter—-in this case, what one is capable of doing—by considering reasons which bear on a question with a distinct subject matter—in this case, whether doing the thing in question or treating it as an option would be a shaming, dishonest, lax, or grotty thing to do. We saw that it was a crucial part of Moran's explanation that this would be a legitimate thing to do only if one had a kind of first-person authority here, where one has first-person authority insofar as one's attitude is determined by one's answer to the *deliberative question*, by one's *deliberative commitment*. One has first-person authority over one's beliefs insofar as one's belief is determined by one's answer to the deliberative question of what to believe. Similarly, then, one has first-person authority over one's deliberative options insofar as one's deliberative options are determined by one's answer to the deliberative question of whether to treat some option as an option, and to answer the deliberative question one must reflect on certain features of one's action, whether it would be a shaming, dishonest, lax, or grotty thing to do.

If this is correct, then we have an argument for the striking conclusion that we have such first-person authority over our deliberative options—what we can and cannot do—when these are moral capacities and incapacities are partly rationally determined by our deliberative commitments. If we take Williams' observations about knowledge of moral incapacity as facts which a theory of knowledge of moral incapacity must explain, and we take Moran's theory to provide the best explanation of this observation, and that theory requires us to assume that we have a kind of first-person authority over our deliberative options, then we have an argument for the conclusion that we have such authority over what we can and cannot do.

4.3 An Objection and a Clarification

I want to end this discussion of knowledge of moral incapacity by briefly discussing a common line of objection to transparency based theories like Moran's. Addressing this objection will allow me to clarify the central notion of a deliberative commitment in the theory. The response to this objection developed here will then be applied to a structurally similar objection raised to Williams's theory of moral incapacities later in the paper.

The line of objection concerns cases of self-knowledge where it is not plausible to claim that one engages in deliberation on outward directed questions and cases where one has some attitude but not for any particular reason.[15] It may seem as though Moran's account cannot explain how we arrive at self-knowledge in such cases. Moran has responded to this line of objection, and I think that his response is basically correct.[16] It is true that Moran's theory is best motivated by examples where there is some deliberation about the outward directed question, but it should not be assumed that such deliberation is essential to the theory. What is essential is taking the deliberative stance towards a question about one's own attitudes: one lets one's answer to this question be fixed by one's answer to the deliberative question. Now, in many cases, one will have to "make up one's mind" on the deliberative question before one can arrive at an answer to the theoretical question. But in other cases, one's mind will already be made up on the deliberative question. And in other cases still, the deliberative question will be answered one way or another and it doesn't even make sense to speak of making up one's mind with respect to the question. What matters is that one's mind *is* made up on the deliberative question, that the deliberative question is answered one way or another, that one *has* a deliberative commitment. It is *this* which allows one to answer the theoretical question by taking a deliberative stance towards it. It is the lack of a settled answer to the deliberative question which explains cases where one is alienated from one's attitude, and cannot answer the theoretical question about it by taking a deliberative stance towards the question. In many cases where one's mind is already made up on some deliberative question, it will not make sense to speak of one's mind being made up for any particular reason. The question may be so central to one's overall system of attitudes that the deliberative question simply cannot come up. As Moran insists, in such cases, it may nonetheless be true that what one's attitudes are is determined, insofar as one is rational, by one's answer to

---

[15] This line of objection can be found in (Shoemaker 2003), (Shah and Velleman 2005), and (Byrne 2005).

[16] Moran's responses can be found in (Moran 2003) and (Moran 2012).

the corresponding deliberative questions, by one's deliberative commitments.[17]

5 A Theory of Moral Incapacities

Having argued that our knowledge of moral incapacities has the features Williams claims it has, and having sketched a theory which vindicates these claims, we have completed our case for (3)—the claim that first-personal ascriptions of moral incapacities can be arrived at on the basis of practical deliberation, in a way that is both immediate and exhibits a kind of transparency. The question now is whether we can provide a theory of moral incapacities on which they can be known by this distinctive means *and* which vindicates (1) and (2)—the claim that third-personal ascriptions are ascriptions of genuine incapacities and the claim that third-personal ascriptions and first-personal ascriptions are about the same subject matter respectively. I will now argue that we can.

Let us begin by asking: what must moral incapacities be such that we are able to know them on the basis of practical deliberation—that is, on the basis of considering reasons for and against performing particular actions, or for and against treating some option as an option in deliberation? In light of the conclusion of the previous section, it seems that any answer to this question will have to accommodate the fact that moral incapacities are the kind of thing we have a certain degree of first-person authority over. In this section, I present Williams's own theory of moral incapacities and argue that his theory does not adequately explain how it is possible to know our moral incapacities on the basis of practical deliberation. I then offer a theory of moral incapacities which improves on Williams's in this respect, drawing on elements of the discussion of the previous sections.

Williams himself takes the task of giving a theory of moral incapacities to be that of distinguishing moral incapacities from other kinds of incapacities, including physical incapacities and other psychological incapacities. Williams begins this task by noting that "A moral incapacity belongs to the species: incapacity to do a certain thing knowingly" (Williams 1995, 48). According to Williams, this distinguishes moral incapacities from physical incapacities and many psychological incapacities. However, this is not enough to distinguish moral incapacities from certain kinds of merely psychological incapacities. As Williams says: "[o]ther incapacities...take the form of one's being unable to ϕ if one knows that one is ϕ-ing" (Williams 1995, 49). Williams goes on: "These might include such things as walking a narrow plank over the Avon gorge; having a relaxed and

---

[17] For this response, see especially (Moran 2012, 219–24). See also (Hieronymi 2006, 2009) on the relationship between settling questions and holding attitudes.

15

friendly conversation with the lover of one's spouse; and eating a roast rat" (Williams 1995, 49). So Williams then considers the suggestion that the cases might be distinguished in terms of trying. In the case of moral incapacity, there is a sense in which if I were to try, I might succeed, whereas in the cases of mere psychological incapacity, if I were to try I will not succeed. Disgust, for instance, will prevent me from eating the roast rat. The suggestion is that "moral incapacity consists in the agent's not being able (even) to try" (Williams 1995, 49). But, of course, this will not do, since in the other cases the agent may not even be able to try: disgust, for instance, may prevent me from even trying to eat the roast rat.[18]

It is at this point that Williams makes the central move of his discussion. He writes that "To understand moral incapacity, we have to consider more closely the way in which the incapacity is connected with the agent's reasons" (Williams 1995, 50). Williams claims that, "there is a difference between moral incapacity and the others, which is to be found in the way in which the agent's incapacity is connected to the mode in which the action presents itself to him; there is a corresponding difference in the way in which he discovers that he cannot ϕ 'through' imagining ϕ-ing" (Williams 1995, 50). So there are two differences to be found here. The first difference, according to Williams, is that, in the case of a moral incapacity, "the fact that an act would be (in my view) disloyal or shabby is a consideration *for me* in deciding not to do it". He notes that the fact that eating a roast rat would be disgusting *could* be a reason for him, but when it explains his incapacity, it doesn't explain it in the same way. I take Williams's point here to be that, in the case of eating a roast rat, the fact that it would be disgusting may, along with knowledge I have about myself, lead me to conclude that I cannot eat the roast rat. And then "the question of deciding not to do it will not come up". Or, perhaps if I nonetheless think that it is something I can do, and I do decide to try, "the incapacity will make me fail". It is not something I can do knowingly. So, in these cases, the fact that it is disgusting is a reason, but not a decisive reason *for me* in deciding not to do it even though I might discover that I cannot do it through imagining. In contrast, in the case of moral incapacity, the consideration is a reason for me, and the way I discover that I cannot do it through imagining is different: "In the case of moral incapacity, my deliberative conclusion not to do the act, reached on the basis of these totally decisive considerations, just is the conclusion that I cannot do it" (Williams 1995, 51).

---

[18] See also (Williams 1981, 129–30).

Now, what do these remarks amount to? Well, they do not amount to a *theory* of moral incapacity of the sort that we are looking for. They only further specify the differences in how an agent arrives at conclusions about what they can and cannot do. These remarks simply explore in more detail the observation that we can arrive at knowledge of moral incapacities by considering reasons which bear on deliberative questions, like the question of what to do, or whether to treat some option as an option in deliberation. Other psychological incapacities cannot be known this way. One finds out that one cannot eat roast rat on the basis of the kind of evidence that one would use in finding out that someone else cannot eat roast rat.

Williams, does, however, go on to make some suggestive remarks about the connection between practical deliberation and moral incapacities. In particular, he writes that "the idea of a possible deliberation by the agent in such terms gives us the best picture of what the incapacity is" and "[w]e understand the agent's moral incapacity just because we understand how 'I can't' could be the conclusion of his deliberation" (Williams 1995, 51). But we need to understand not only how 'I can't' could be the conclusion of practical deliberation, but how it could be a way of coming to *know* that one cannot do the thing in question. We face the two-topics problem: how could it be *legitimate* to reach such a conclusion on the basis of practical deliberation? And we have seen that an answer to this question must appeal to something like first-person authority. At best, we get from Williams's discussion the claim that certain dispositions of the agent are the grounds of the agent's incapacities, and that these dispositions are, roughly, dispositions to treat certain considerations as grounds for concluding that one cannot do the thing in question, they are dispositions to deliberate in certain ways. Thus Williams writes:

> [T]he dispositions that are the ground of the agent's incapacities are focussed on to a particular case through the way in which the features of the case impress themselves on the agent—ways that are best represented by a deliberation, even though the deliberation need not consciously occur. (Williams 1995, 52)

But to be told that there are certain dispositions to reason or deliberate in certain ways which are the grounds of an agent's incapacity is only to be told something very coarse-grained about moral incapacities. True, it may suffice to distinguish moral incapacities from other kinds of incapacities—so it may suffice for Williams's purposes. But it leaves us without an answer to the two-topics problem.

It is here that I think we can do better than Williams and offer a more fine-grained theory of moral incapacity, one which solves the puzzle about knowledge of moral incapacities, and which relates moral incapacities to deliberation in the way that Williams wants to. To do so, we need to clarify the nature of the dispositions involved in arriving at the deliberative conclusion. We can better understand these dispositions by seeing them as dispositions to treat certain considerations as resolving particular deliberative questions and to thereby arrive at a kind of deliberative commitment on the matter—that is, a commitment to a particular answer to the deliberative question. The deliberative question for the agent is whether to treat some option as an option in deliberation. Can I include my student's idea in my essay unacknowledged? Am I to even treat it as an option, something to include in my deliberation about what to do? Is the fact that it would be a shabby thing to do a reason for me to not even consider it? In answering such deliberative questions, the agent arrives at a decision or commitment of some sort. I decide not to treat including my student's idea in my essay unacknowledged as an option, I arrive at a deliberative commitment to not treat it as an option in my deliberation about what to do. So we must see the dispositions involved as dispositions of deliberative reflection and as closely tied to the notion of deliberative questions. And in doing so we see that the dispositions involved in explaining moral incapacities are just like those dispositions involved in answering deliberative questions about our actions and attitudes more broadly.

We are now in a position to see that it is not so much the relation to deliberation here that matters but an agent's answer to particular deliberative questions about which options to treat as options and which considerations to treat as reasons for ruling certain options out. An agent may be committed to ruling some option out, or treating some consideration as a reason for ruling some option out without having explicitly deliberated about it. When the question of whether she can do the thing in question arises for her, she might simply avow an answer which is in line with her commitments or she may engage in deliberation about whether to treat some option as an option or whether to treat some consideration as a reason. Thus we capture the sense in which "deliberation need not consciously occur". And here we see how moral incapacities can be tied to "the sort of person the agent is", to their character. For what an agent is committed to treating as a deliberative option and as a reason for treating an option as an option tells us a lot about the kind of person the agent is. As Williams says, it tells us a lot about a person that they would even treat an option as an option. Someone who does not treat the suffering of others as a reason for not even treating options which would lead to the suffering of

others as an option, and, say, weighs the pros and cons of the option, may already appear diminished in our eyes.

The crucial element in this explanation is the notion of an agent's commitment to a particular answer to a deliberative question, the notion of a deliberative commitment. This element is what is missing in Williams's account. To speak merely of dispositions to deliberate is to obscure the fact that what underlies these dispositions is a kind of deliberative commitment, a commitment which reflects the agent's view of the reasons for and against treating some option as an option in deliberation. It also leaves it somewhat obscure as to how these dispositions might ground knowledge of moral incapacities. By appealing instead to deliberative commitments, which are elements in the explanation of other forms of self-knowledge, we can help ourselves to a unified explanation of knowledge of moral incapacity and other kinds of self-knowledge. Insofar as our deliberative commitments have a kind of authority over the elements of our psychic life they concern, we can be said to have first-person authority over those elements, and can come to know those elements by reflection on reasons bearing on having them. It is because deliberative commitments on the truth of particular propositions have authority over what we believe that we can come to know what we believe by considering evidence bearing on the truth of the propositions believed. Likewise, it is because deliberative commitments on the good, useful, and the obligatory, have authority over which options are available to us that we can come to know what we can and cannot do by considering questions about the good, the useful, and the obligatory.

Now, I want to end my case for this theory of moral incapacities by noting how it can be appealed to explain a central feature of moral incapacities. An early statement of the idea of a moral incapacity occurs in Williams' "Critique of Utilitarianism". He writes:

> It could be a feature of a man's moral outlook that he regarded certain courses of action as unthinkable, in the sense that he would not entertain the idea of doing them: and the witness to that might, in many cases, be that they simply would not come into his head. Entertaining certain alternatives, regarding them indeed as *alternatives*, is itself something that he regards as dishonourable or morally absurd. (Smart and Williams 1973, 92)

We want a theory of moral incapacities which can explain this feature: that certain options will simply not come into the agent's head, that an agent may simply not regard them as options. We can explain this in terms of an agent's deliberative commitments,

where the more central of these can be thought of as partly constitutive of their moral character. A man's "moral outlook" will be partly constituted by his deliberative commitments, which concern which options he is willing to treat as options in deliberation. Insofar as he is committed to not treating certain options as options in deliberation certain courses of action simply won't come into his head and will not be treated as serious options in his deliberation or if they do, he will immediately recognise that the option is of a kind that he is generally committed to not treating as a deliberative option.

6 Objections and Alternatives

In this final section, I will examine two objections and alternatives to the theory of moral incapacities I have just defended. I begin by addressing an objection to the role played by reason in the theory of moral incapacity just offered. I then address the objection of the skeptic who holds that the theory given cannot vindicate the claim that third-personal ascriptions and third-personal ascriptions of moral incapacities are about the same subject matter.

6.1 *Brute Incapacities*

Craig Taylor has argued that the case of Huck Finn raises a challenge to theories of moral incapacities like Williams's. Huck Finn, you will recall, is incapable of turning his friend Jim in to the authorities. This appears to be a paradigmatic example of a moral incapacity—Huck's incapacity has its source in his moral agency or character. Taylor argues that this example raises a problem for Williams's account. For, Williams appears to be committed to holding that Huck must be disposed to reason to the conclusion "I cannot turn Jim in" from reasons that he has for not turning Jim in. That is, on Williams's view, Huck's moral incapacity must be grounded in such a disposition. Yet, Taylor argues, it is implausible to think that Huck has such a disposition. Taylor wants to suggest that the case of Huck Finn is a case where the incapacity is "a primitive fact about him" and "cannot be explained in terms of some deliberation over his dispositions and commitments" (Taylor 2001, 59). Why should we think this? Taylor asks us to consider the justification-seeking why-question "Why can't you turn Jim in?" directed at Huck. He suggests that answers like "I promised to help Jim" and "Jim has been a good friend to me" or even "It is the only decent thing" will not do, for "the point of the story is that for Huck such considerations do not even count here as reasons for action (or inaction)—not, that is to say, in the case of such a slave" (Taylor 2001, 60–61). This, indeed, is the standard

20

philosophical upshot of this case: Huck does not think of these considerations *as* reasons (Bennett 1974). Since Huck does not think of these considerations *as* reasons, he simply cannot be disposed to reason from them to the conclusion that he cannot turn Jim in, according to Taylor. His incapacity, therefore, cannot be explained in terms of such a disposition.

On the face of it, this example raises a serious problem for theories like the theory offered in the previous section. We seem to have an example where the agent has a moral incapacity but is not disposed to deliberate in the way that Williams claims they must be, and who lacks the kind of deliberative commitment we have appealed to in our version of the theory. How might we respond? There are two lines of response we might develop.

The first line of response challenges Taylor's reasoning about the example. Taylor argues from the assumption that Huck does not think of some consideration *as* a reason to the conclusion that the consideration cannot *be* Huck's reason for reaching the relevant deliberative conclusion. But we may reject this line of reasoning by pointing out that there are many cases where an agent draws a conclusion, and acts, for a reason which they do not see *as* a reason. There is a general distinction between what an agent believes to be a reason in the normative sense, and an agent's reasons for acting (Alvarez 2010). Huck's reason for not turning Jim in may well be that Jim is his friend *even though* Huck does not think of the fact that Jim is his friend *as* a reason, in the normative sense, for not turning him in. This is not to say that the fact that Jim is his friend is merely an explanatory reason for Huck's not turning Jim in—it may genuinely be *his* reason for not turning Jim in, in the sense of being an agent's reason. That an agent's reasons and an agent's beliefs about normative reasons can come apart like this has been convincingly argued by Kieran Setiya (Setiya 2007, 36–38). So, if we take this line of response, then we may reject Taylor's premise and hold that Huck is, nonetheless, disposed to deliberate as Williams claims even though he does not believe that the consideration from which he may conclude that he cannot do the thing in question is a reason against doing it. The plausibility of this line of response comes down to how plausible it is to think that Huck might reply to the justification seeking why-question with an answer like the following: "The reason why I cannot turn Jim in is that he is my friend, even though I do not think that this is a reason for not turning him in at all, and think that I in fact have every reason to turn him in!".

The second line of response is to question the assumption that the deliberative conclusion must always be drawn from a distinct premise at all. Williams's way of presenting the

deliberative dispositions certainly suggests as much, but we might resist the suggestion. As we saw earlier, in response to a common line of objection to transparency theories of self-knowledge, it is possible to hold that someone might have a deliberative commitment, that is, they may be committed to a particular answer to a deliberative question, for no particular reason. The commitment may be so central to one's overall system of attitudes and commitments that it does not make sense to speak of it being held for a particular reason. Indeed, it must often be the case that one is committed to the truth of particular propositions for no particular reason at all. I may be committed to the truth of the proposition that I was born, although I do not believe this for any particular reason. And, as Anscombe has argued, it must sometimes be the case that we act intentionally for no reason at all, where the correct answer to the justification seeking why-question is "no reason" (Anscombe 1963, 25–28). But in both cases, there is an important difference between being deliberatively committed to some belief or action for no reason, and not being so committed. The former brings with it a kind of answerability or responsibility to reason: insofar as one is so committed one exercises a kind of rational authority over what one believes and what one does which one would not if one were not so committed. Were new reasons bearing on your belief or action to come to light, you would be responsible for revising your commitment. According to this line of response then, Taylor's reasoning gives us no grounds for doubting that Huck is not deliberatively committed to rejecting the option of turning Jim in. We have no grounds for thinking that the deliberative question of whether to treat turning Jim in as an option, is not negatively resolved for Huck. Indeed, it is quite plausible that many cases of moral incapacities involve the mere recognition that a particular option falls under a general description where one already has a deliberative commitment against treating options which fall under that description as options in deliberation. In such cases the agent may not have any further reason for not treating the option as an option, and the deliberation that takes place in resolving the specific deliberative question may take a very attenuated form.[19]

_____

[19] A similar thing can be said in response to the objection to Moran's transparency theory that there is no deliberation involved in answering the question "Do you believe that there are bicycles on Mars?". Insofar as one is committed to a negative answer to the question of their being bicycles anywhere else but on Earth, one simple applies one's more general deliberative commitment to the specific deliberative question about whether there are bicycles on Mars. Notice that in such a case it would be hard to provide a particular reason for believing that there are no bicycles on Mars in answer to the question "Why do you believe that there are no bicycles on Mars?" None of this does anything to show that one is not deliberatively committed to a negative answer to the deliberative question of whether one is to believe that there are no bicycles on Mars. It just shows how little cognitive effort is involved in a deliberative question being resolved for one and how it doesn't always make sense to request a reason for belief in cases of deliberative commitment.

That one or the other of these lines of response (or both) must be on the right track can be brought out by raising a challenge for Taylor's alternative. Whatever else we say about Huck's case, it still seems plausible to say that Huck will be in a position to know that he cannot turn Jim in in a way that does not require him to infer this from evidence about himself. So our claim about the epistemology of moral incapacity holds good in this case, whatever we think about Huck's reasons and dispositions to deliberate. But if this is so, then we should be motivated to treat this case as like other cases where the agent does have a reason for ruling the relevant option out. And we can appeal here to the common factor of the agent having resolved the deliberative question, having reached a deliberative commitment here, in order to explain their ability to know their moral incapacities in both cases. In claiming that Huck's incapacity is a brute psychological fact about him, Taylor leaves himself with no way of explaining the distinctive means by which Huck can come to know that he cannot turn Jim in.

6.2 *A Special Sense of 'Cannot'*

I noted earlier that there is a temptation to argue from Claim 3, the claim about first-person ascriptions of moral incapacities to either the falsity of (1) or (2). The basic idea here is that the sense of 'cannot' involved in the first-person deliberative conclusion either expresses a kind of deontic modality or a special kind of circumstantial modality tied to rationality and the weight of reasons—'I can't rationally do it, the reasons against it are too strong'.[20]

I will argue that although Williams's argument against this kind of scepticism is inconclusive, there is an argument available to show that the third-person and first-person senses are the same and that they do not express a special kind of circumstantial modality. I will then argue that the theory of moral incapacities I have defended can vindicate this claim.

In his discussion of this kind of scepticism, Williams assumes that the sceptic must be claiming that the special sense of 'cannot' involved in the first-person deliberative conclusion expresses a kind of deontic modality. His case against the sceptic, so conceived, is strong. He first makes the point about the implications of third-person ascriptions and then, rather enigmatically writes: "'Can't' is, once again, 'can't', and it has the same sort of implications as it does in the third-person cases for the question whether the world

---

[20] Daniel Dennett has suggested that moral incapacities involve this kind of modality (Dennett (1984) p. 133). See (Watson 2002, 139–41) for discussion

will contain my doing the thing in question" (Williams 1995, 48). But this is something that the sceptic who holds that the first-personal 'I can't' expresses a special kind of circumstantial modality can concede. They might hold that the sense of 'cannot' which 'signifies whatever rejection is embodied in the agent's deliberation' *does* express something an agent cannot literally do, but in a distinct circumstantial sense of 'cannot', one tied to rationality and the weight of reasons.

I suspect that this is exactly the kind of position held by those who are committed to what Williams calls the "boring answer" to the question of what moral incapacities are and what its co-relative, practical necessity, is. Williams writes:

> That question might have had only a rather boring answer: for instance, that *must* is selected when the preferred course of action is very markedly favoured over the others, or the weight of reasons overwhelmingly comes down on one side. (Williams 1981, 126)

A similar point applies in the case of moral incapacities. Perhaps 'I cannot' is selected when the weight of reasons overwhelmingly comes down against the option in question. There may well be a circumstantial sense of 'cannot' corresponding to such cases. We needn't think that the 'I cannot' is deontic. Perhaps there is a sense in which the agent is unable to do the thing in question when the reasons are so strongly against it. Why couldn't the boring answer be right in such a case?

It is at this point that we can draw on considerations from our discussion of knowledge of moral incapacity in presenting an argument against the boring answer and the view that the sense of 'cannot' involved in first-person ascriptions of moral incapacities is tied to cases were one has overwhelming reasons against some option. As Williams points out, in cases like these "a set of objectives or constraints is merely taken for granted, and relative to them, a particular course of action is very clearly singled out" (Williams 1981, 126). What this suggests is that in such cases the deliberative question the agent faces is this: "What shall I do?", where the deliberative options are taken as fixed and she is to choose between them. Now, the sceptic we are arguing against agrees that the first-personal ascription 'I cannot' is arrived at on the basis of deliberation, that is, by a distinctive means. If so, then we can think of the agent as arriving at the conclusion 'I cannot' by answering the deliberative question of what she is to do. Now here is the problem for this suggestion. It is by answering the deliberative question of what she is to do that the agent answers the question of what she *will* or *will not* do. We saw this suggestion earlier

in the discussion of knowledge of action. What would entitle her to the stronger conclusion in this case, namely that she *cannot* do the thing in question, when the deliberative question she is answering is whether *to do* the thing in question? Answering *this* deliberative question, and arriving at a corresponding commitment, *at best* supports the conclusion that she *will not* do the thing in question. In order to arrive at a stronger conclusion she must answer a different deliberative question. But then if the deliberative question is whether to treat some option as an option, then she is no longer deliberating against a background of options she takes for granted and weighing reasons for and against them, so the conclusion she reaches cannot merely be one which reflects the fact that she cannot do the thing in question because she has overwhelming reasons against doing it, it must be one which reflects the fact that she has taken some reason as a reason for ruling some option out, for answering the deliberative question of whether to treat that option as an option, and has thereby arrived at a deliberative commitment of sorts.

If this argument is correct, then even if there is a sense of 'cannot' which corresponds to an agent being unable to do something in the face of overwhelming reasons, this is not the sense of 'cannot' involved in first-personal ascriptions of moral incapacities, and the sceptic has no grounds for objecting to the claim that first-personal and third-personal ascriptions of moral incapacities are about the same subject matter.

Bibliography

Alvarez, Maria. 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford University Press.

Anscombe, Elizabeth. 1963. *Intention*. Harvard University Press.

Ashwell, Lauren. 2013. "Deep, Dark ...or Transparent? Knowing Our Desires." *Philosophical Studies* 165: 245–56.

Bar-On, Dorit. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford University Press UK.

Bennett, Jonathan. 1974. "The Conscience of Huckleberry Finn." *Philosophy* 49: 123–34.

Boyle, Matthew. 2009. "Two Kinds of Self-Knowledge." *Philosophy and Phenomenological Research* LXXVIII (1).

———. 2011a. "'Making up Your Mind' and the Activity of Reason." *Philosopher's Imprint* 11 (17): 1–24.

———. 2011b. "Transparent Self-Knowledge." *Proceedings of the Aristotelian Society Supplementary Volume* LXXXV.

———. 2019. "Transparency and Reflection." *Canadian Journal of Philosophy*, 1–28.

Byrne, Alex. 2005. "Introspection." *Philosophical Topics* 33 (1).

———. 2011a. "Knowing What i Want." In *Consciousness and the Self: New Essays*, edited by JeeLoo Liu and John Perry. Cambridge University Press.

———. 2011b. "Transparency, Belief, Intention." *Proceedings of the Aristotelian Society Supplementary Volume* LXXXV.

———. 2018. *Transparency and Self-Knowledge*. Oxford University Press.

Calhoun, Cheshire. 2009. "What Good Is Commitment?" *Ethics*, no. 119.

Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.

Cassam, Quassim. 2015. *Self-Knowledge for Humans*. Oxford University Press.

Cox, Ryan. 2018. "Knowing Why." *Mind & Language* 33 (2): 177–97.

Dennett, Daniel. 1984. *Elbow Room*. Cambridge, Mass.: MIT Press.

Edgley, Roy. 1969. *Reason in Theory and Practice*. Hutchinson & Co.

Evans, Gareth. 1982. *The Varieties of Reference*. Edited by John McDowell. Oxford University Press.

Fernandez, Jordi. 2013. *Transparent Minds*. Oxford University Press.

Finkelstein, David H. 2003. *Expression and the Inner*. Harvard University Press.

Frankfurt, Harry. 1998a. "Rationality and the Unthinkable." In *The Imporance of What We Care about*. Cambridge University Press.

———. 1998b. "The Importance of What We Care About." In *The Imporance of What We Care about*. Cambridge University Press.

———. 1999. "On the Necessity of Ideals." In *Necessity, Volition, and Love*. Cambridge University Press.

———. 2002. "Reply to Richard Moran." In *The Conours of Agency: Essays on Themes from Harry Frankfurt*. The MIT Press.

Gaita, Raymond. 2004. *Good and Evil: An Absolute Conception*. Abingdon, Oxon; New York, NY: Routledge.

Gertler, Brie. 2011. "Self-Knowledge and the Transparency of Belief." In *Self-Knowledge*, edited by Anthony Hatzimoysis. Oxford University Press.

Hampshire, Stuart. 1975. *Freedom of the Individual*. Chatto & Windus.

Hampshire, Stuart, and H. L. A. Hart. 1958. "Decision, Intention and Certainty." *Mind* LXVII (265): 1–12.

Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophy Quarterly* 87 (1): 45–74.

———. 2009. "Two Kinds of Agency." In *Mental Actions*, 138–62. Oxford University Press.

Keeling, Sophie. 2019a. "Knowing Our Reasons: Distinctive Self-Knowledge of Why We Hold Our Attitudes and Perform Actions." *Philosophy and Phenomenological Research*.

———. 2019b. "The Transparency Method and Knowing Our Reasons." *Analysis* 79 (4).

Kloosterboer, Naomi. 2015. "Transparent Emotions? A Critical Analysis of Moran's Transparency Claim." *Philosophical Explorations* 18 (2).

Lawlor, Krista. 2009. "Knowing What One Wants." *Philosophy and Phenomenological Research* 79: 47–75.

McGeer, Victoria. 1996. "Is "Self-Knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation." *Journal of Philosophy* 93 (10): 483–515.

———. 2008. "The Moral Development of First-Person Authority." *European Journal of Philosophy* 16 (1).

Moran, Richard. 2001. *Authority and Estrangement*. Princeton University Press.

———. 2003. "Responses to o'brien and Shoemaker." *European Journal of Philosophy* 11 (3): 402–19.

———. 2004. "Précis of Authority and Estrangement: An Essay on Self-Knowledge." *Philosophy and Phenomenological Research* LXIX (2).

———. 2012. "Self-Knowledge, 'Transparency,' and the Forms of Activity." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press.

Paul, Sarah K. 2014. "The Transparency of Mind." *Philosophy Compass* 9 (5): 295–303.

Setiya, Kieran. 2007. *Reasons Without Rationalism*. Princeton University Press.

———. 2011. "Knowledge of Intention." In *Essays on Anscombe's Intention*. Harvard University Press.

Shah, Nishi, and J. David Velleman. 2005. "Doxastic Deliberation." *The Philosophical Review* 114 (4).

Shoemaker, Sydney. 2003. "Moran on Self-Knowledge." *European Journal of Philosophy* 11 (3).

Smart, J. J. C., and B. Williams. 1973. *Utilitarianism: For and Against*. Cambridge University Press.

Taylor, Craig. 1995. "Moral Incapacity." *Philosophy* 70.

———. 2001. "Moral Incapacity and Huckleberry Finn." *Ratio* XIV.

Van den Beld, T. 1997. "Moral Incapacities." *Philosophy* 72 (282).

Watson, Gary. 2002. "Volitional Necessities." In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 129–59. The MIT Press.

Way, Jonathan. 2007. "Self-Knowledge and the Limits of Transparency." *Analysis* 16: 223–30.

Williams, Bernard. 1981. "Practical Necessity." In *Moral Luck*, 124–31. Cambridge University Press.

———. 1995. "Moral Incapacity." In *Making Sense of Humanity and Other Philosophical Papers*. Cambridge University Press.