BERNARD WILLIAMS'S INTERNALISM: A NEW INTERPRETATION

By

Micah J Baize

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson John Bricke

_____

Dale Dorsey

_____

Ben Eggleston

_____

Richard De George

_____

Maria Carlson

Date Defended: April 11, 2012

The Dissertation Committee for Micah J Baize
certifies that this is the approved version of the following dissertation:

BERNARD WILLIAMS'S INTERNALISM: A NEW INTERPRETATION

_____

Chairperson John Bricke

Date approved: April 11, 2012

Abstract

There has been significant and continued debate over the nature and truth of Bernard Williams's internalism. My aim is to resolve much of the dispute over both of those issues by providing a new interpretation of his internalism—the reasons^{H} interpretation. To explain the new interpretation I make a distinction between there being a reason to perform an action (a reason^{E}) and an agent having a reason to perform an action (a reason^{H}). For an agent to have a reason to perform an action, it must be within the agent's capacity to perform the action for that reason. According to the reasons^{H} interpretation, internalism is the claim that in order for an agent to *have* a reason, it must be within the agent's capacity to be motivated to perform the action. An important consequence of this interpretation is that externalists with respect to the previous interpretations can consistently accept the truth of internalism on the reasons^{H} interpretation.

To support the accuracy of this new interpretation of Williams's internalism, in Chapter 1 I argue that the predominant interpretations are problematic because they inconsistent with one of two claims which are most likely essential to a correct interpretation. In Chapter 2 I then provide a detailed explanation of the reasons^{H} interpretation as well as three considerations which together strongly support the plausibility of it as a correct interpretation. Chapter 3 completes the argument that the reasons^{H} interpretation is the most charitable interpretation with respect to Williams's argument against external reasons. In Chapter 4 I defend the truth of internalism against various objections that have been raised against the doctrine. Lastly, in Chapter 5 I will show that the same concern which underlies Williams's explanation and defense of internalism is the same concern which is the basis for his rejection of the "morality system"—a particular conception of morality which he addresses in *Ethics and the Limits of Philosophy*.

Acknowledgments

Completing a dissertation cannot be done without the help of many people along the way. I am very appreciative of all of the work that Jack Bricke has provided with the dissertation as well as my philosophical education as a whole. I am also grateful for the work of my other dissertation committee members. Ben Eggleston has served on every committee related to my graduate work. Dale Dorsey, Richard De George, and Maria Carlson served on both my dissertation committee and my comprehensive exam committee. The comments and questions of all four improved my understanding of my own work and its relation to the larger body of philosophical thought. Thanks are also due to Cindi Hodges for the secretarial help she provided throughout the dissertation writing process. Lastly, a special thanks goes to my wife Emily for her constant encouragement and support as I wrote the dissertation as well as in life.

Table of Contents

Preface

In this dissertation my objective is to provide a new interpretation of Bernard Williams's internalism. There has been significant and continued debate over the nature and truth of internalism. My aim is to resolve much of the dispute over both of those issues. Both issues will be largely resolved by providing a new interpretation of Williams—what I call the reasons^{H} interpretation.

Internalism is related to the issue of reasons for action. One of the central questions related to the subject matter of reasons for action is whether, for there to be a reason for action an agent must have some motivation related to the action. There are two broad positions with respect to this issue. The first position, subjectivism, claims (amongst other things) that there is a normative reason for an agent to perform an action only if he has a motivation to perform it. The other position, objectivism, claims that there can be a normative reason for an agent to perform an action even if there is not a related motivation. For example, if an agent has no motivation to pursue a college education, subjectivists would deny that there is a normative reason for the agent to pursue one, while objectivists would claim that it could still be the case that there is a normative reason for the agent to pursue it.

In "Internal and External Reasons", the first of several articles he wrote in defense of internalism, Williams claims that in order for there to be a normative reason for an agent to perform an action, there must be some motivation related to the putative reason for action.[1] This claim has led many to think that Williams's internalism either presupposes or is a defense of subjectivism. In what follows I will argue for a different understanding. In part, my claim is that

[1] Bernard Williams, "Internal and External Reasons", in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 101.

Williams is relying on a different conception of "normative reason" than most readers have suspected. The notion of "normative reason" in the debate between subjectivists and objectivists is roughly the concept of what features of a situation *count in favor of* performing an action. Williams's notion of "normative reason", if I am correct, is instead the concept of a consideration which an agent is actually capable of acting upon. So, if there is to be a normative reason for an agent to perform an action (in this sense), it must be that the agent is actually capable of performing the action. So, when Williams claims that in order for there to a normative reason for an agent to perform an action the agent must have some motivational element related to the putative reason, he is claiming that, in order for the agent to be capable of performing an action for the putative reason, the agent must have the motivational capacity to be motivated by the putative reason to perform the action. In the example above, if the agent truly did not have the motivational capacity to pursue a college education then there is not a normative reason for the agent to pursue it in Williams's sense of "normative"—because he is not capable of doing so. For the sake of clarity, the former type of reasons will be labeled "reasons that there are", and the latter will be "reasons that an agent has".

Importantly, Williams's use of the latter conception of normative reason is not intended to deny the legitimacy of the former conception. On my interpretation, internalism *allows* that there could be a normative reason to perform an action in the former sense even if a particular agent does not have the motivation necessary to perform the action (and so there is not a normative reason in the latter sense). That is because internalism is concerned only with the latter conception. So, if I am correct, internalism does not have any direct impact on the subjectivism/objectivism debate—since it is concerned with a different notion of "normative reason". That my interpretation of Williams's internalism entails that it is compatible with both

subjectivism and objectivism will strike many as a rather implausible claim. It would then seem to be a fairly non-controversial position. Some might even say it is trivial. And, it will likely be objected, surely Williams would not have spent so much writing in defense of a trivial claim. Although that is a legitimate concern, over the course of the dissertation I will provide several arguments in favor of the new interpretation which I think outweigh it.

The plan for the dissertation is as follows. In the first chapter I will aim to show that the current interpretations of internalism are problematic in that they are incompatible with one of two claims which (I will argue) are most likely essential to a correct interpretation of internalism. Given the problems with them, there will be good reason to consider a new interpretation of internalism. In Chapter 2 I will provide a detailed explanation of the reasons^{H} interpretation. I will also provide three considerations which together strongly support the plausibility of it as a correct interpretation. In Chapter 3 I will then show that it is the most plausible interpretation because it is the most charitable interpretation of Williams's argument against external reasons. In addition, I will respond to several potential objections to the accuracy of the reasons^{H} interpretation. If the above arguments and responses are successful, there will be then be very good reason to accept the reasons^{H} interpretation as the correct interpretation of Williams's internalism.

In Chapter 4 I will argue for the truth of internalism. To do so, I will provide some considerations in its favor. Most importantly, however, I will defend its truth against various objections that have been raised against the doctrine. For the most part, what will be shown is that those objections rest upon incorrect understandings of the nature of internalism, and so are not problematic for the correct interpretation (the reasons^{H} interpretation). Lastly, in Chapter 5 I will show that the concern which underlies Williams's explanation and defense of internalism is

the same concern which is the basis for his rejection of the "morality system"—a particular conception of morality which he addresses in *Ethics and the Limits of Philosophy*.

Chapter 1: Problems with Current Interpretations of Williams's Internalism

The overall objective in this chapter is to explain why the predominant interpretations of Williams's internalism are unsatisfactory—that is, to explain why they are likely not accurate interpretations of Williams's internalism. That fact will give us good reason to consider the possibility of an alternative interpretation of Williams. In this chapter I will first explain that interpreters of Williams have overlooked a distinction between two different types of reasons for action. They have failed to make a distinction between the *reasons that there are* for an agent to act, and the *reasons that an agent has* for acting.[1] (That failure is due in large part to Williams's lack of clarity on the issue.) This distinction will be explained shortly. But, because of this mistake, interpreters have taken internalism to be a claim about the reasons that there are, and not the reasons that an agent *has*. I will argue that it is actually a claim about the latter.

To show that there is something amiss with the current interpretations of internalism, I will identify two claims which a correct interpretation of internalism will most likely need to uphold. Neither of the predominant interpretations of internalism will prove consistent with both of those claims. That conclusion will prepare us for the second chapter, in which I will explicate the new interpretation of Williams as well as begin the defense of the accuracy of that interpretation.

I. Reasons that there are vs. reasons that an agent has

[1] Alan Goldman makes a similar *terminological* distinction in his "Reasons Internalism", *Philosophy and Phenomenological Research* 71, no. 3 (2005). However, it relies upon a desire-dependent conception of practical reasons and is substantively different than my account. Unlike my account, it denies that there can be a reason^{E} (a concept which will be explained in what follows) for action if an agent has no motivation relevant to the reason.

When we consider what reasons for action there are for a particular agent, there are at least two different concerns we can have. One, we can be concerned with what there is reason to do *irrespective of the agent's limitations*. We might imagine what there would be reason for an agent to do *if* he had unlimited powers, knowledge, and so on. On the other hand, we could be concerned with what there is reason for the agent to do *given the agent's limitations*. Since agents can have limitations which constrain what actions they can actually perform, we can consider what there is reason to do within those limitations. There might be good reason for me to play professional baseball, because it would be fun, provide plenty of opportunities to travel, and possibly provide an opportunity to play in a World Series, but since I do not have the physical abilities necessary to play on a professional team, I don't *have* a reason to play pro baseball. In order to avoid confusion in discussing these two different types of reasons, we will make a terminological distinction between them.

Reasons that there are. If we are concerned with what there is reason for an agent to do irrespective of the agent's limitations, we are concerned with the reasons that there are. What reasons there are to act in a situation depends upon the features of the situation that count in favor of performing an action. The features that count in favor of performing an action will depend on the correct account of practical reason. Some accounts of practical reason take desires to generate reasons, some take value, some pure reason, and others some combination of these (and perhaps others). But, whichever account is correct, that feature would determine what reasons there are to act. And when we are concerned with the reasons that there are to perform an action, we are not concerned with whether the agent is able to actually perform it. If going to see a movie would generate a reason to do so (either because an agent desired to do so, or it would promote the value of pleasure, etc.), there is a reason even if an agent is unable to go to the

movie (perhaps the movie theater is clear across town and he does not have transportation to arrive in time for the movie). There is a reason to watch the movie; it is just a reason which cannot be acted upon.

For the sake of clarity, in the text I will indicate that we are concerned with this kind of reason by using the notation "reason^{E}" (a reason that exists).

Reasons that an agent has. Amongst the reasons that there are, only a subset of them are ones that a particular agent is capable of acting upon. Due to various limitations (e.g. physical, temporal, etc.) agents cannot act on some, if not most, reasons that there are. Only those reasons^{E} which a particular agent can act on are ones that the agent has. A reason to act which is one that an agent is capable of acting upon is one that an agent *has*. It could be that there is a reason to swim across the Atlantic Ocean. Perhaps my desire to do so or the good that would be promoted by doing so generates a reason to swim across it. However, because I do not have the physical capacity to do so (I would surely drown within the first few miles), I do not *have* a reason to swim across the Atlantic. To have a reason to perform an action, an agent must be capable of performing the action.

In addition to physical limitations, it is almost certain that human agents also have psychological limitations. They have psychological incapacities which prevent them from being motivated to perform some actions. And, it is not just that they merely chose to perform some alternative action; it is that they *could not* have performed the action in question at all. Suppose that there is a reason^{E} for people to go to the top of the Sears Tower to see the view from that height. Even if there is this reason^{E}, someone might have such a fear of heights that he could not bring himself to go to the top of the Sears Tower. If so, then he has a psychological limitation which makes it impossible for him to be motivated to go to the top of the Sears Tower. And, if

limitations which prevent us from performing an action preclude us from *having* a reason to do so, then it seems that the agent in question does not have a reason to go to the top of the tower.

This last claim risks begging the question with respect to the truth of internalism as it is to be understood on my interpretation, since on my interpretation internalism is just the claim that a psychological inability to be motivated to φ precludes an agent from having a reason to φ. In the fourth chapter I will provide arguments for its truth.

But for now I want to help clarify the notion of "having a reason" by making a few points. The first issue is that to have a reason φ for reason *r*, it is not sufficient to merely have some motivational element related to *r*. To have a reason to φ requires that the whole of one's motivational set be properly related to *r* such that one could be motivated by *r* to φ. That the latter is required is due to the effect of the interplay between the motivational elements in an agent's S. Although an agent may have some motivation to φ for *r*, it may be that he has other, stronger, motivations which prevent the former motivation from being effectual.

The second point is that it is not enough that one be capable of being motivated by *r* when it is considered by itself. Instead, one must be capable of being motivated by *r*, given all other reasons. Although an agent might be able to be motivated by a reason *r* to φ when *r* is considered in itself, given other reasons and actions available to him, he may not be able to be motivated by *r* to φ *all things considered*. If so, then *r* is not a reason he has to φ.

The reason for both of the above points can be seen in the following example. Suppose that a good friend of mine is getting married and has invited me to attend the wedding. Because he is a good friend, and I think that his wedding is an important event in his life, I am motivated to go to the wedding. However, I am also a huge Kansas City Royals fan, and the day after I was invited to the wedding, I am given tickets to a World Series game, which the Royals are playing

in, and which happens to be at the same time as the wedding. Because my devotion to the Royals is so great, attending the game takes precedence over any other possible activities. As a result, I could not bring myself to attend the wedding. I have some motivations related to attending the wedding. And, I could even be motivated to attend the wedding, taking into consideration my entire motivational set—but only if I do not also consider the fact that I have tickets to the World Series. But in determining whether I have a reason to attend the wedding, the question is whether, given the actual circumstances, I am capable of being motivated to attend the wedding. Once all things are considered, in particular all of the elements of my motivational set and the opportunity to attend the Royals' World Series game, I am not actually capable of being motivated to attend the wedding. For that reason, I do not *have* a reason to go to the wedding—since it is not something I am capable of doing.

A third point about having a reason is that, in order for an agent to have a reason to act, it must be the case that there is also a reason^{E} to perform the action. Reasons^{H} are a subset of reasons^{E}. If there is not a reason^{E}, then an agent cannot have a reason to perform it.

We might think that an agent can have a reason to act even if there is not a reason^{E}. Most likely this will be a scenario where an agent has a false belief which we think he is rational to act upon. For example, if someone falsely believes that there is a knock at the door, we might think that he has a reason to open the door, even though there is not a genuine reason^{E} to do so. (There is not a reason to open the door because there is no one to let in.) If we think the agent has a reason to open the door, that is only because we are confusing reasons^{H} with the reasons that an agent is *epistemically rational* in acting upon. The latter is determined by what the agent believes and what it makes sense to do in light of those beliefs. When we say that an agent has a reason to

act, we are not saying that it makes sense in light of his beliefs. Rather, we are saying that there is a reason^{E} for him to perform an action and that he is capable of doing so.

It seems fairly clear that Williams's internalism is *not* a theory about what actions it is epistemically rational for an agent perform. He states explicitly, and multiple times, that if an agent's decision to perform an action is based on a false belief, then he does not have a reason to perform that action.[2] Therefore, Williams is almost certainly not concerned with what there is reason from the agent's epistemic perspective to do. That is, he is not concerned with giving a theory about an agent's *epistemic* reasons for action.

One last point about reasons^{H} is important. To say that an agent *has* a reason is not (in the context of this discussion about reasons for action) the same as saying that the agent *possesses* the reason. There is a use of "has a reason", especially with respect to reasons for belief, which might suggest such an understanding. To say that an agent has a reason to believe *p*, we often mean that the agent is in possession of some evidence which gives him a reason to believe *p*. However, if that were what we meant by claiming that an agent has a reason to act, the claim that reasons^{H} to act are constrained by the limitations of an agent to act on them would likely be false, at least absent additional argument. Consider reasons for belief. An agent can have a reason to believe *p* even if he is incapable of believing it. Suppose that Jones saw Smith steal a book. However, because Jones is a life-long friend of Smith's and has a high regard for his character, he cannot bring himself to believe that Smith actually stole the book. Since he has evidence for Smith's having stolen the book, Jones possesses a reason to believe that Smith stole it—and so he has a reason to believe it—but it is not within his capacity to do so. So, in this case, an

[2] Williams, "Internal", 102-3. Bernard Williams, "Internal Reasons and the Obscurity of Blame", in *Making Sense of Humanity*, (Cambridge: Cambridge University Press, 1995), 36. Bernard Williams, "Postscript: Some Further Notes on Internal and External Reasons", In *Varieties of Practical Reasoning*, ed. Elijah Millgram (Cambridge: The MIT Press, 2001), 91.

incapacity does not preclude the agent from having a reason. Therefore, if we were to use "has a reason" in the same way with respect to reasons for action, the claim I made earlier that the reasons an agent has are constrained by the limitations of the agent would be suspect. The other reason not to understand "has a reason" in the possession sense is that it does not seem to reflect Williams's use of the phrase.[3] This will be seen over the course of the next few chapters.

With the distinction between "is a reason" and "has a reason" in hand, my interpretation of Williams's internalism can be more clearly stated. According to my interpretation of internalism, when Williams says that all reasons are internal, he means that all reasons an agent has to act upon are internal.[4] More particularly, in order for an agent to have a reason to φ, it must be possible for the agent to be motivated to φ for that reason. The notion of "possibility" in use here has not been clear. As will be argued for later, "possible" should be understood to mean "within an agent's capacity".[5] So, in order for an agent to have a reason to φ, it must be within his capacity to φ for that reason. Internalism is merely the claim that psychological limitations—an inability to be motivated by a reason—are constraints on the reasons an agent has most, just as physical limitations are. By interpreting internalism to be a claim about the reasons that an agent has, as opposed to the reasons that there are, many (if not all) of the concerns of those who currently reject internalism will be dissolved.

---

[3] One of Williams's uses of the phrase (on page 104 of "Internal and External Reasons") might seem to suggest the possessive notion of the phrase. However, his use of the phrase is open to alternative interpretations, and, given the whole corpus of his work, we have good reason not to accept the possessive use.

[4] This reflects his claim in the "Postscript". However, much more must be argued to establish that I am actually interpreting him correctly, and so that he is using the phrase "has most reason" in the same way as I am here.

[5] The example I have given here may be thought to beg the question in favor of internalism (on my interpretation of it). Since my claim is that internalism is the thesis that psychological limitations nullify an agent having most reason to φ, it might appear that I am building into the notion of "having a reason" the requirement that an agent must be psychologically capable of acting upon it. There are two things to be said in response. One, the reason I used a psychological limitation is that it does not appear that physical capacities can have the same type of effect on other physical capacities. It is not clear how an agent who is physically capable of acting upon a reason when considered in itself would not be physically capable of acting upon the reason when all reasons are considered. Second, and more importantly, if someone rejects the claim that psychological limitations preclude an agent having most reason to φ, it is open to them to dismiss this example. As well, it seems that the distinction between having a reason and having most reason would become moot.

Given the distinctions I have made between different kinds of reasons, it should be pointed out that in explaining and defending internalism, Williams often uses the various phrases ("reason for action," "there is a reason", "has a reason," etc.) without specifying the type of reason with which he is concerned. At times he seems to use them interchangeably; at other times not. The confusion over internalism is due in part to this. Since there is not a consistent use of terminology, or at least it is not made explicit, we will have to determine which type of reason he is concerned with by evaluating his claims about reasons for action. In chapter two I will argue that the claims he makes about reasons for action are most plausible if they are about reasons^{H}. In fact, interpreting them as being about reasons^{E} would be an uncharitable interpretation of Williams.

But, since we have assigned specific meanings to the phrases "is a reason" and "has a reason" it could be misleading to use either of these phrases when discussing Williams. However, in discussing reasons for action, he often uses the phrase "normative reason".[6] The claim "there is a normative reason for an agent to φ" is ambiguous as to whether we are claiming that "there is a reason for the agent to φ" or "the agent has a reason to φ". In order to minimize confusion as well as avoid begging the question in favor of either "there is a reason" or "has a reason", when discussing Williams's claims about reasons for action I will often use the phrase "there is a normative reason" (except when directly quoting him).

II. Two claims likely essential to a correct interpretation of internalism

[6] Williams use of "normativity" has been the point of some contention. In Chapter 2 I will point out that there are two types of different types normativity and I will also argue that philosophers have misunderstood the notion of normativity which Williams is using.

I have two objectives in the rest of this chapter. First, I want to explicate the predominant interpretations of internalism (which exclude mine). One important aspect of the explications will be to explain the rationale for each interpretation. The second objective is to give at least a strong *pro tanto* reason to reject each as an accurate interpretation of Williams. The interpretations considered will contradict one of two claims of Williams about internalism, claims which I will argue are essential to his position. Since the interpretations conflict with the claims, we will have at least a strong *pro tanto* reason for rejecting them.

The two claims which I think are essential to Williams's internalist position are the following.

(R) All reasons for action are relative to an agent's subjective motivational set.

(N) No particular conception of practical reason is presupposed by internalism.

A correct interpretation of Williams will most likely incorporate both of these claims, unless reasonable possibilities for doing so are exhausted. An interpretation which can uphold these two claims is, *ceteris paribus*, to be preferred over those which cannot. In this chapter I will aim to show that each of the predominant interpretations of internalism, which take internalism to be concerned with reasons^{E}, are inconsistent with one of these claims. That they are inconsistent does not prove that the interpretation is necessarily unacceptable. It may be that other considerations will force us to conclude that an interpretation which conflicts with one of these claims is still to be preferred. If so, that would indicate that Williams was just mistaken about an aspect of internalism.

1. (R) All reasons for action are relative to an agent's subjective motivational set

There are good reasons for thinking that the correct interpretation of Williams will most likely be consistent with (R). Williams repeatedly states that the reasons for action of agents are relative to their S. In "Internal and External Reasons" (IER) he writes that, "basically, and by definition, any model for the internal interpretation must display a relativity of the reason statement to the agent's subjective motivational set…"[7] This seems to indicate the relativism claim is an essential, non-negotiable, feature of internalism. In addition, at the end of IER he claims that we cannot "define notions of rationality where the action rational for *A* is in no way relative to *A*'s existing motivations".[8] In "Internal Reasons and the Obscurity of Blame" (IROB) Williams restates the internalist claim that "A has a reason to φ only if he could reach the conclusion to φ by a sound deliberative route from the motivations he already has."[9] And in his reply to John McDowell's critique of internalism, Williams states that "[i]t follows [from internalism] that what an agent has a reason to do will be a function of…the existing set of his motivational states."[10] The general idea in these quotes is that an agent's normative reasons for action must have some relationship to the agent's current motivations.

Although we might be skeptical of Williams's claims about the relativity of reasons for action, we have to keep in mind here that we are only concerned with identifying the correct *interpretation* of Williams's internalism, and not with assessing whether his view is correct. We should not be concerned with the truth of any aspect of internalism, except insofar as its implausibility would give us good reason to reject it as a charitable interpretation of Williams.

[7] Williams, "Internal", 102.
[8] Ibid., 112.
[9], 35.
[10] Williams, "Replies", in *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, ed. J.E.J. Altham and Ross Harrison (Cambridge: Cambridge University Press, 1995), 187.

Given both the abundance of Williams's claims about the relativity of reasons and the fact that we are at present only concerned with identifying a correct interpretation of Williams, it is highly likely that the relativity claim is essential to the internalist position.

2. (N) No particular conception of practical reason is presupposed by internalism

Next, what about (N), the claim that internalism does not presuppose a theory of practical reason? Why should we resist an interpretation which makes internalism presuppose a theory of practical reason? There are two considerations which support the idea that a correct interpretation of Williams's internalism will likely be consistent with (N). But, I first need to explain the distinction between a theory of practical *reason* and a theory of practical reason*ing*.

A theory of practical reason is an account of what features of a situation count in favor of an agent performing an action. There are two predominant general views on this matter.[11] The first view is *subjectivism*, which states that what generates a reason to act is an agent's desires. The fact that an agent desires something constitutes a reason for the agent to perform an action which is relevant to satisfying that desire. There are of course nuances to this view. Most subjectivists think that not just any desire generates a reason for action. Instead, it is only those desires which an agent would have, were the agent to have full-information. Since the exact details of this qualification are largely irrelevant to our discussion, we will not examine them further here. The second view is *objectivism.* According to it, what generates a reason for action is not the desires of the agent. As Sobel puts it, any view which claims that the generators of reason to act are "not to be found in the agent's contingent proattitudes" counts as an objectivist

[11] Here I am following David Sobel's use of terminology in his "Subjective Accounts of Reasons for Action", *Ethics* 111, no. 3 (2001).

theory.[12] Objectivism is then a very broad category, intended to include all but the subjectivists. However, I shall add one qualification to this, one which perhaps makes my characterization of the dichotomy vary to some degree from Sobel's. Objectivists can allow (though some will deny this) that desires generate *some* reasons to act. However, what they must deny is that desires are the *sole* generators of reasons to act.

A theory of practical reason*ing*[13] however, is not (at least primarily) concerned with what features of a situation generate reasons to act; instead it is concerned with what type(s) of practical deliberation it is appropriate for an agent to engage in when trying to determine what action to perform. Such a theory could be concerned with what considerations the agent should think about and what principles of practical reason he should rely on in coming to his practical decisions. It will also very likely take into account the agent's epistemic situation in assessing what it is rational for the agent to do. Due to an agent's false beliefs, either about the situation that he is in or perhaps also about what principles of practical reason are correct, it may not be epistemically rational for the agent to perform an action which there is most reason (according to the theory of practical reason) for the agent to perform. The theory of practical reason*ing* might conclude that there is most reason *from the agent's epistemic perspective* to perform an alternative action.

To see how a theory of practical reason*ing* may come to a different conclusion than a theory of practical *reason* about what action there is reason to perform, consider the following case. Suppose that subjectivism is true and that I have a desire for some lemonade. If some neighborhood children have a lemonade stand just down the street, then, according to a theory of

[12] Sobel, "Subjective", 474.
[13] It might also be described as a theory of practical *rationality*. However, to use this phrase specifically for this type of theory would be confusing in the context of the discussion of Williams's internalism, as he says that internalism is concerned with an agent's rationality, but his use of that term has a different meaning.

practical *reason*, there would be a reason to go outside and buy a lemonade from them, as drinking the lemonade would satisfy my desire. However, although there is a reason to do so, I am unaware that the children are outside. I do not have any information which would lead me to suspect that they are outside (perhaps they have never had a lemonade stand before). Despite my ignorance, a subjectivist theory would still say that there is a reason to go outside and purchase the lemonade, because doing so would satisfy a desire. But, a theory of practical reason*ing* would give a very different answer. A theory of practical reason*ing* is likely going to take into account my epistemic condition and would therefore conclude that I instead have most reason to drive to the gas station to buy a lemonade (of course, if I do leave the house to drive to the gas station and see the kids, given the change in my epistemic situation I may *at that time* come to have an epistemic reason to buy the lemonade from the neighborhood kids).

We are now ready for the two considerations that support the claim that a correct interpretation of Williams's internalism will likely be consistent with (N). The first is that Williams never claims that his theory rests upon any particular theory of practical *reason*. Granted, this consideration amounts only to an argument from silence. Unfortunately (from my perspective), Williams never explicitly states "Internalism does not presuppose any particular theory of practical reason." However, given that Williams wrote several articles on internalism, were *he* to have taken it to rely upon any particular theory of practical reason, it is reasonable to expect that he would have mentioned that fact in at least one of them (and how *he* understands it is what is important here, since we are concerned with a correct interpretation of *his* theory). Since he did not, there is reason to think that he did not take internalism to presuppose a particular theory of practical reason.

The second consideration is that Williams explicitly states that internalism is very open with respect to what counts as *sound* deliberation. In other words, it is very liberal with respect to what counts as *sound* practical reason*ing*. Williams was familiar with objections to internalism which claimed that it presupposed instrumentalism. His response was not to admit that it did, but instead to reiterate the liberality of the internalist position with respect to what counts as sound rational deliberation. In "Internal Reasons and the Obscurity of Blame" Williams notes that internalism does not restrict practical deliberation to mere means-end reasoning. Instead, he allows for all of the following types of practical deliberation (a list which he does not claim to be exhaustive): "finding a specific form for a project that has been adopted in unspecific terms", the "invention of alternatives", "to think of another line of conduct altogether, as when someone succeeds in breaking out of a dilemma", "the perception of unexpected similarities". He then goes on to add, "[s]ince there are *many ways of deliberative thinking,* it is not fully determinate in general, even for a given agent at a given time, what may count as 'a sound deliberative route'" (emphasis mine).[14]

In "Values, Reasons, and Persuasion", a later writing of Williams, he stresses again the openness of internalism with respect to practical reasoning. He writes:

> As we have seen, the internalist account is generous with what counts as a sound deliberative route. It rejects the picture by which a determinate and fixed set of preferences is expressed simply in terms of its decision-theoretical rational extensions, and deliberation is construed simply as discovering what these are.[15]

[14] Williams, "Internal Reasons", 38.
[15] Williams, "Values, Reasons, and the Theory of Persuasion" in *Philosophy as a Humanistic Discipline,* ed. A.W. Moore (Princeton: Princeton University Press, 2006), 114.

Practical reasoning is not merely determining how to best fulfill one's preferences (though of course that may be part of it). In addition, Williams allows that types of reasoning typically (at least) *thought* to be external *could* be a necessary part of rational deliberation. Moral or prudential considerations are not excluded as possible aspects of rational deliberation. They may be. For example, he allows that a Kantian conception of pure practical reason as defended by Korsgaard could qualify as a proper form of deliberation. However, he requires that before we conclude that moral or prudential considerations are a necessary part of rational deliberation, we must establish the truth of that claim by argument.[16]

What makes the above quotes so important is the fact that they are all given in the context of Williams's explanation of internalism's position with respect to what counts as a *sound* deliberative route. If Williams took internalism to rely upon a particular theory of practical *reason*, then it seems reasonable to expect that he would think it restricts what types of practical reason*ing* are *sound*. For example, *if* he took internalism to presuppose subjectivism, then he should have ruled out practical reason*ing* which is based on non-subjectivist accounts of practical reason. How could he think that practical reasoning which is based on (what he would take to be) a false theory of practical reason would be *sound* reasoning? Since he did not restrict what counts as sound deliberation in accordance with any theory of practical reason (except for it not being based on false information), there is good reason to think that he did not take internalism to presuppose any particular theory of practical reason.

[16] This paragraph on what counts as proper rational deliberation is likely misleading with respect to Williams's position (on my interpretation of it anyway). It might appear that if a Kantian conception is true and it would say that the correct action to perform is φ, then therefore the agent in question necessarily has a reason to φ. On my interpretation that is not necessarily true. Whether the agent has a reason depends upon whether he can also be motivated to perform the action. So, on my interpretation, what Williams is allowing is that agent's practical deliberation can be in a form entirely other than instrumental, but, if the agent is going to have a reason to perform the action which that deliberation requires (perhaps *suggests*?), it must be that he can be motivated to perform it. If not, then he does not have a reason to perform it—although there is a reason^{E} for him to do so.

That the correct interpretation of internalism should most likely be consistent with (N) is perhaps a bit more controversial than the claim that it should be consistent with (R). Many philosophers have rejected internalism precisely because they interpret it as presupposing a particular theory of practical reason—namely, an instrumental (or at least quasi-instrumental) one. But again, I am not claiming that an interpretation according to which it presupposes a theory of practical reason is *necessarily* incorrect. Rather, I am claiming that there is good reason to resist such an interpretation, absent countervailing considerations. Ultimately the correct interpretation *may* end up presupposing a theory of practical reason, but if that is the case, it must be because, all things considered, we cannot find a better alternative interpretation. If Williams's theory actually relies upon a theory of practical reason, he was oblivious to that fact. Though not impossible, it gives us good reason to reject such an interpretation if we reasonably can.

And if it does turn out that internalism does not presuppose any particular theory of practical reason (which is the case if my interpretation of internalism is correct), neither subjectivists nor objectivists should have any qualms about accepting internalism. Holders of either position can accept that an agent's inability to be motivated to act on a genuine reason that there is can nullify that reason from being one that the agent has. To say that an agent must be capable of being motivated to φ for reason *r* in order for him to have a reason to φ does not entail that the motivation to φ is what generates a reason *r* to act, or that the lack of an ability to be motivated to φ for *r* precludes *r* from being a reason to φ.

III. The two predominant interpretations of internalism

For the purposes of my dissertation and my argument for an alternative interpretation of internalism, the predominant interpretations of internalism can be categorized into one of two general interpretations. Amongst these two interpretations there are a variety of sub-interpretations, some more different than others. However, they all share one of two features which I will argue gives us a *pro tanto* reason to reject them. The first (and most common) interpretation is that internalism is the claim that what reasons^{E} there are for an agent to act is relative to the motivations of the agent. The second interpretation of internalism is that it is a necessary condition on the reasons^{E} of a fully rational agent—that is, a reason^{E} for action must be capable of motivating a fully rational agent. If a fully rational agent would not be motivated by a putative reason, it is not a reason to act.

Earlier I argued that a correct interpretation of internalism is most likely going to be consistent with both (R) and (N). The problem with the two interpretations above is that the first is inconsistent with (N), while the second is inconsistent with (R). A consideration crucial to my argument is that the two interpretations take internalism to be a claim about reasons^{E}. It is because they so interpret internalism that ultimately they are unsatisfactory. Any interpretation of internalism that understands it to be concerned with reasons^{E} will not be able to uphold both claims coherently. That is why I think that the reasons^{H} interpretation is to be preferred.

We will now take a look at both of the reasons^{E} interpretations in turn. In doing so, we will be primarily concerned with the rationales for each interpretation, and why it is that each conflicts with either (R) or (N).

1. Reasons^{E} to act are constrained by the subjective motivational set of the agent.

According to this interpretation of internalism, there can be a reason^{E} for an agent to act only if there is an element in the agent's motivational set related to that reason. Internalism is then the thesis that a necessary condition for a reason^{E} is that an agent have the motivation to act on that reason. On this interpretation, Williams's claim that all reasons are relative to an agent's subjective motivational set is taken to be an essential aspect of internalism. (As we will see, the second interpretation disagrees.) However, there are two different understandings of how Williams comes to the conclusion that all reasons are relative. Both of them agree that his claim depends upon a particular theory of practical reason, but they disagree over what that theory is. We will look at both.

A. Internalism presupposes a quasi-instrumental conception of practical reason.

According to this view, internalism's claim that all reasons for action are relative to an agent's S follows only because Williams (whether knowingly or not) presupposes an instrumental or quasi-instrumental conception of practical reason. "Quasi-instrumental" is included here because Williams has denied that his theory relies upon a narrow instrumental conception. Whereas a purely instrumental theory entails that an agent can only reason about how to satisfy the desires he already has, Williams allows for much broader deliberation. In addition to desires, an agent's "dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects…embodying commitments of the agent"[17] are included as part of the agent's motivational set. These are all permitted to influence the agent's reasoning.

This interpretation, let us call it the *quasi-instrumental interpretation*, is by far the most common understanding of Williams's internalism. Of those who accept this interpretation, some accept internalism and some reject it. Russ Shafer-Landau, Brad Hooker, Jay Wallace, and

[17] Williams, "Internal", 105.

Rachel Cohon, amongst others, all reject internalism.[18] That is due, at least in part, because they reject quasi-instrumentalism. Brad Hooker states that "The dispute between Williams and the external reasons theorist is ultimately over the starting points of practical deliberation."[19] Similarly, Shafer-Landau says that, "For Williams, rational deliberation must be rooted in one's existing motivations. On his account, motivation that is deliberatively unrelated to one's already existing motivations cannot be rational."[20] Externalists (on this interpretation of internalism) think that there can be reasons^{E} to act, even if it is not related to the agent's motivations.

However, not all philosophers who interpret internalism in this manner reject internalism. Understandably, instrumentalists most likely will not have a problem with it. Identifying those who accept internalism on this interpretation is a bit more difficult, but Alan Goldman is one.[21]

Leaving aside the issue of who accepts it and who does not, we must now ask why philosophers have arrived at this interpretation. We can begin by first noting the Humean-like nature of internalism. Williams maintains there is a close, *though not identical*, relationship between his view and Hume's.[22] Hume is of course well known for claiming that reason is a slave to the passions. Reason does not tell us what we ought to do. Rather, reason is merely instrumental in determining how to satisfy our desires. Since Williams denies that that there is an identical relationship with Hume's view, we have to be careful to determine what aspects of Hume's view it shares and which ones it does not. The most certain difference is Williams's

[18] Their interpretations can be found in the following. Russ Shafer-Landau, *Moral Realism*, (Oxford: Clarendon, 2003). Brad Hooker, "Williams' Argument against External Reasons," *Analysis* 47, no. 1 (1987): 42-44; Jay Wallace "Three Conceptions of Rational Agency, *Ethical Theory and Moral Practice* 2 (1999): 217–242; Rachel Cohon, "Internalism about Reasons for Action," *Pacific Philosophical Quarterly 7*4, no. 4 (1993): 265-288. Others who accept this interpretation: Derek Parfit, "Reasons and Motivation", *Proceedings of the Aristotelian Society, Supplementary Volumes* 71 (1997): 99-146; John Robertson, "Internalism about Moral Reasons," *Pacific Philosophical Quarterly* 67 (1986): 124-135; Joshua Gert, "Williams on Reasons and Rationality," in *Reading Bernard Williams*, ed. Daniel Callcut (New York: Routledge, 2009): 73-93; Elijah Millgram, "Williams's Argument against External Reasons," *Noûs*, 30, no. 2 (1996): 197-220.
[19] Hooker, *Williams' Argument*, 44.
[20] Shafer-Landau, *Moral*, 172.
[21] Goldman, "Reasons Internalism".
[22] Williams, "Internal," 102.

much broader conception of practical deliberation, which allows for deliberation beyond a narrow instrumentalism and which at least allows for (though it does not necessarily accept) a Kantian conception of pure practical reason. Such a conception of practical reason Hume would certainly deny.

We might think that such an allowance would appease any would-be externalists. However, despite that divergence from Hume, Williams still insists upon the Humean-like nature of internalism. To see why, we must consider a couple statements Williams makes about what we can say about an agent's reasons for action. First, he claims that, in saying what there is reason for an agent to do, we are allowed to go beyond what the agent is currently motivated to do. That is, there can be a reason for an agent to φ even if he is not presently motivated to φ. Second, we are allowed to correct for "any errors of fact and reasoning involved in the agent's view of the matter."[23] So, if an agent lacks a motivation to φ, but this is due to a false belief or to poor reasoning, then there could still be a reason to φ.

Given that an agent can have a reason to perform an action which he is not currently motivated to do, that we can correct for the agent's beliefs and reasoning, *and* that Williams allows for the possibility of non-Humean practical reasoning, we might think internalism does not imply that reasons are relative after all. For example, if we are objectivists about practical reason, what there is reason to do will be determined in large part independently of an agent's desires (though perhaps not entirely). And objectivists are likely to think that there is a reason for an unemployed person to get a job (because it would be prudential for him to do so). It seems then that we can say there is a reason for the person to get a job even if he lacks the motivation to do so. If the agent is not motivated by the reason to get a job, objectivists would claim that the unmotivated agent not reasoning properly. Were he reasoning properly, he would be motivated

[23] Williams, "Internal Reasons", 36.

to get a job. Therefore, since internalism allows us to correct for an agent's reasoning, it might seem that internalism would allow us to say that there is a reason for the agent to get the job.

However, Williams denies this. He writes:

> The internalist proposal *sticks with its Humean origins* to the extent of making correction of fact and reasoning part of the notion of 'a sound deliberative route to this act' but not, from outside, prudential and moral considerations. To the extent that the agent already has prudential or moral considerations in his S, of course, they will be involved in what he has reason to do. (italics mine)[24]

In other words, we can ascribe a reason^{E} for action to an agent only if he already has motivations relevant to it. If the person does not already have the motivation to get a job (or at least some motivation relevant to it, such as the motivation to pay his rent), then there is not a reason^{E} to get a job. If the agent completely lacks any motivation related to getting a job, then there is not a reason^{E} for the agent to do so.

What is key here for understanding Williams (on this interpretation) is that, although we are allowed to correct an agent's beliefs and reasoning in determining what there is reason^{E} to do, to do so there must *already* be some motivational element in the agent's S related to the action. A comparison between two similar scenarios can highlight what is going on here. Suppose that two different people share similar circumstances. Both of them have no motivation to look for a job. Suppose also that, unbeknownst to them, they would both get the next job they would apply for. Is there a reason^{E} for each person to apply? According to internalism, that depends.

[24] Williams, "Internal Reasons," 36-7.

Suppose that after looking for a job for the last year with no success, and with news reports of a worsening economy, Bob loses all motivation to look for a job. He has no desire to apply to any more job postings. Suppose also that Charles has never had any motivation to look for a job. And he has this lack of motivation, despite knowing that without a job he will lose his rental and will have to move back in with his parents. Unbeknownst to both Bob and Charles, however, were they each to apply to the next job they hear about, they would be hired. Given this fact, can we say that there is a reason^{E} for them to apply?

According to internalism, there is a reason^{E} for Bob, but not for Charles. Bob's lack of motivation is due to incorrect information. Were he to know that he would be hired, he would be motivated to apply for the job. Therefore, we can say that there is a reason^{E} for him. But not so for Charles. And that is because, even if he knew that he would get a job, he would still lack the motivation. He enjoys not working, and he prefers to continue to be unemployed, even if it means moving back in with his parents. Since he would continue to lack any motivation to get a job even if he had correct information about whether he would get hired, there is no reason^{E} for him to look for a job.

Most objectivists would object to this latter conclusion of internalism. Surely there is *a* reason^{E} to look for a job. It does not even have to be overriding. Perhaps Charles lives in very comfortable circumstances and so, although there is a reason to look for a job, the reasons to not look for a job outweigh the reason to look for one. But, if Charles truly has no motivation to get a job, then there is not a reason for him to apply for one.

It should be clear now why some would interpret Williams to be relying upon at least a quasi-instrumental conception of practical reason. It seems to imply, if not explicitly state, that all reasons^{E} depend upon an agent having a motivation relevant to the reason. This makes

reasons^{E} to act motivation-dependent. A reason^{E} cannot exist unless an agent has a motivation related to it. So, internalism does seem to rely upon some type of quasi-instrumental conception of practical reason.

The quasi-instrumental presupposition of internalism can best be seen in a further claim of Williams. According to Williams, even an agent's *needs* do not generate a reason^{E} for an agent to act unless the needs are related to the agent's motivations. It is understandable to think that, even if some types of considerations do not generate reasons absent an agent's motivations, surely an agent's genuine needs would do so. But Williams also denies this. He states:

> I take it that insofar as there are determinately recognisable needs, there can be an agent who lacks any interest in getting what he indeed needs. I take it, further, that that lack of interest can remain after deliberation, and, also that it would be wrong to say that such a lack of interest must always rest on false belief.[25]

So, on Williams's view, even the genuine needs of an agent do not generate reasons^{E} to act in and of themselves. There can be a reason to fulfill one's needs only if the agent is motivated to do so (were the agent to have correct information).

Externalists (on this interpretation of internalism) think, however, that there can be reasons to act even if they are not related to our motivations. Most externalists support their position by imagining a situation in which it appears obvious that there is a reason^{E} for an agent to act—and where this is so even if the agent is not motivated to act upon it. As an example, consider an argument against internalism given by Shafer-Landau.[26] He asks us to imagine

[25] Williams, "Internal," 105.
[26] Shafer-Landau, *Moral Realism*, 185-7.

someone who suffers from melancholy and so, although she (let us call her Debbie) can imagine what it would be like to have a social life, has no motivation to perform the actions necessary for having one. She has no desire to join a social club, a book-discussion group, or even to call an old friend. Doing these things would make her life go better. And so they give her a reason to do them, even if she lacks the motivation to do them. We can represent Shafer-Landau's argument as follows.

1. If internalism is true, then there is not a reason for Debbie to develop a social life if she is unmotivated to develop one.

2. There is a reason Debbie to develop a social life even if she is unmotivated to develop one.

C. Hence, internalism is false.

Whether we will accept Shafer-Landau's argument depends largely on our conception of practical reason. But, for that reason it is successful in showing that the truth of internalism (on the interpretation of Shafer-Landau, et al.) depends heavily upon what theory of practical reason is true. Since this interpretation makes the truth of internalism rely upon an instrumental or quasi-instrumental theory of practical reason, it is inconsistent with (N). Hence, that gives us a *pro tanto* reason to reject the Shafer-Landau, et al., interpretation.

B. Internalism depends upon a novel conception of a reason for action. We now turn to the other view of internalism which also interprets it as the claim that reasons^{E} to act are constrained by the subjective motivational set of the agent. Although this view agrees with the first about the claim of internalism, it diverges from it in its understanding of the basis for that

claim. On this view, the relativism claim does *not* depend upon instrumentalism or quasi-instrumentalism. Instead, Stephen Finlay argues that internalism rests upon a novel conception of "reason for action". He claims that Williams's argument "begins from a substantive and interesting analysis of the *concept* of a normative reason."[27] And, if we accept Williams's conception of a reason for action, then all reasons for action are internal reasons—all reasons for action are relative to the agent's motivations. Let us call this interpretation the *novel conception interpretation.*

According to Finlay, Williams's conception of a "reason for action" is "*R* is a reason for *A* to *φ*' means that *R is an explanation of why A would be motivated to φ if he deliberated soundly*."[28] He also says that Williams's conception is heavily influenced by Davidson's "Actions, Reasons, and Causes". Finlay bolsters this claim by pointing out the historical context within which "Internal and External Reasons" was written. At that time, the now common distinction between normative and motivating (or explanatory) reasons was not common.[29] Finlay writes that Davidson "argued that the concept of "reasons" that justify an action just is a certain concept of "reasons" that causally explain it; a justification is simply a rationalizing explanation in terms of the agent's desires and beliefs."[30] Therefore, when an agent performs an action, the explanation of the desires and beliefs of the agent resulting in the action is the normative reason for the action. However, Finlay clams that Williams is unsatisfied with that unqualified claim. Williams thinks that we should deny that an explanation of an agent's performance of an action constitutes a normative reason, *when the agent acts on the basis of a false belief*. Although it is understandable why they would have performed the action given that

[27] Stephen Finlay, "The Obscurity of Internal Reasons*,"Philosopher's Imprint* 9, no. 7 (2009): 1.
[28] Ibid., 14.
[29] Ibid., 13.
[30] Ibid.

they had the false belief, there was not a *normative* reason for them to perform the action. To rule out such explanations as constituting a normative reason, Finlay claims that Williams provides an improved conception of "reason for action". That conception is that a normative reason for action is an "explanation of an agent's action under the condition of the absence of false belief or ignorance (i.e., "sound deliberation")."[31]

However, Finlay claims that Williams is not concerned primarily with the bare concept of a reason for action. Instead, he is primarily concerned with what it is to be a reason for action *for an agent* to φ.[32] But his concept of a reason for action is important because it impacts what it is to be a reason for action *for the agent.* The latter can only be determined after determining what constitutes the former.

For Williams, in order for a reason for action to be one *for the agent*, the agent must be capable of recognizing the reason. And he must recognize the reason on the basis of true beliefs. That is why Williams claims that there is a reason for the agent to φ only if the agent could recognize the reason through sound deliberation.

Crucial to the idea that the agent must be capable of recognizing the reason is that, to recognize the reason, the agent would have to be able to recognize that he would act on the reason if he were correctly informed. And, whether *he* would act on the reason depends, so Williams argues, on the type of person that he is. And, given that the elements of his S constitute, at least in part, the type of person that he is, whether he would act on the reason depends on the elements of his S. So, if the agent is going to recognize a reason as being one that there is for him to act upon, the agent is going to have to see some type of relationship between the reason to act *r* and his S such that he could see himself being motivated by *r* to act.

[31] Ibid., 14.
[32] Ibid., 15.

Note that the conception of practical reasoning in play here is not an instrumental conception. It is not as though the agent is merely deliberating about how to best satisfy his desires. Rather, he is deliberating about whether a consideration is one which he would be motivated to act upon were he correctly informed. Finlay says that this is an *evidential* model of practical reason.[33] This is to a large degree what separates Finlay's interpretation from the interpretation of Shafer-Landau, et al.

For an example of the evidential model, consider someone who has been told that he should spend the summer as a park ranger. Suppose that the reason given as to why he should is that it would give him time to "discover himself" because there would be ample time spent in solitude. According to Finlay's interpretation, whether that is actually a reason for him to be a park ranger over the summer depends upon whether he could see himself being a park ranger for the reason of discovering himself. And, whether he could see himself doing so is going to depend on the content of his S. If the agent is one who loves the outdoors and is a rather reflective sort of person (these are elements of his S), he could probably see himself being a park ranger for the summer, and for the reason that he could discover himself. In that case, it would be a reason for him to act.

On the other hand, suppose he has a different S. Instead of loving the outdoors and solitude, he loves urban cities with miles of cement, the constant noise of traffic and late-night revelers, and is a people-person who can hardly spend an hour alone. In this case he is very unlikely to see himself as spending a summer as a park ranger for the sake of discovering himself. If he cannot see himself doing that action for that reason, then it is not a reason to act.

Although Finlay acknowledges that this is an unorthodox interpretation of Williams, he thinks that it avoids some serious problems with the other theories. Perhaps most importantly, it

[33] Ibid., 16.

explains why Williams thinks that the reasons there are for an agent is relative to the agent's S, and yet upholds Williams's claim that his theory does not rely solely on an instrumental account of practical reasoning. For Williams, an agent must be able to envision him- or herself acting on a putative reason^{E} in order for it to be a reason^{E}. Agents are not restricted to instrumental reasoning. But, they must see a reason as being one which they would be motivated by in order for it to actually be a reason. And, sources of motivation are not restricted just to the fulfillment of desires. If an agent sees himself as someone who considers more than just how he might fulfill his desires, then the reasons he would see himself acting on would not just be ones aimed at fulfilling a desire. An agent could be motivated because an action would be good, or moral, etc. But, because whether an agent reasons merely instrumentally or not depends on the contents of his S, what reasons an agent sees himself as being motivated by will be constrained by the elements of his S.

Williams's argument for internalism appears to presuppose an instrumental theory of practical reasoning since it claims an agent's reasons for action are relative to the agent's S. However, on this interpretation, the instrumental presupposition is merely apparent, and results from Williams's novel non-instrumental conception of what it is to be a reason for action. Finlay does not say whether we should reject internalism or not. He thinks that because Williams's theory depends on a novel conception of a reason for action, we must first evaluate the plausibility of that conception.[34] Only then would we be in a position to evaluate the truth of internalism.

Although Finlay's interpretation avoids making internalism rely upon an instrumental model of practical reason, it is still problematic because it presupposes a theory of practical reason—the evidential model. It is inconsistent with (N). Again, given that Williams provides

[34] Ibid., 20.

no definite constraints on legitimate types of practical deliberation, we should avoid attributing to internalism a particular theory of practical reason, unless no better interpretation is available.

Those who interpret internalism as the claim that reasons^{E} to act are constrained by the subjective motivational set of the agent have two different understandings for the basis of that claim. The most common one is that it presupposes a quasi-instrumental conception of practical reason. Finlay disagrees, and thinks that it instead rests upon the evidential model of practical reasoning. However, because both of them take internalism to presuppose some theory of practical reason, they are inconsistent with (N). Therefore, we have a *pro tanto* reason to reject these interpretations as correct. We will now take a look at the other main interpretation of internalism.

2. Reasons^{E} to act must be capable of motivating fully rational agents.

On this interpretation, internalism is the claim that reasons^{E} for action must motivate fully rational agents. So, let us call this the *fully rational interpretation.* What is most significant about this interpretation is that it does not take Williams's claim that all reasons for action are relative to an agent's motivations to be essential. This interpretation is shared by both Christine Korsgaard and Michael Smith.[35]

According to Korsgaard, the internalism requirement is that, "Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational

[35] Although Korsgaard and Smith agree on the above, they disagree over whether internalism is a necessary or a necessary and sufficient condition. Korsgaard understands it to be merely a necessary condition, whereas Smith takes it to be a necessary and sufficient condition. However, since this disagreement is relevant only if their more general interpretation is correct, we will leave the question of whether it is a necessary and/or sufficient condition aside.

persons."[36] Or as she puts it elsewhere, "it requires…that rational considerations succeed in motivating us insofar as we are rational."[37] If there is to be a reason^{E} for an agent to φ, it must be that the agent would be motivated to φ *if the agent were fully rational*. If a fully rational agent would not be motivated by the putative reason, then it is not a reason^{E} to act.

Michael Smith expresses a similar understanding of internalism. He writes, internalism is the claim that, "it is desirable for an agent to φ in certain circumstances C, and so she has[38] a reason to φ in C, if and only if, if she were fully rational, she would desire that she φs in C."[39] In order for some putative reason^{E} for action to be an actual one, it must be the case that a fully rational agent would be motivated to perform an action on the basis of it in those particular circumstances.[40]

With this understanding of internalism, Korsgaard and Smith both claim that the relativity of reasons claim does not necessarily follow. If follows *only if* the correct account of practical reason entails that they are relative. Therefore, they reject Williams's claim that "any model for the internal interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*".[41] As they see it, that is merely an erroneous inference by Williams. Williams only thought that internalism entailed that all reasons are relative.
Take as an example of this interpretation Smith's response to Williams's relativistic claim.

[36] Christine Korsgaard, "Skepticism about Practical Reason", *The Journal of Philosophy* 83, no. 1 (1986): 11.
[37] Ibid., 15.
[38] Although Smith uses the phrase "has a reason," he is using it in the sense of "there is a reason" as stipulated above.
[39] Michael Smith, "Internal Reasons", *Philosophy and Phenomenological Research* 55, no. 1 (1995): 112.
[40] The "particular circumstances" qualification is intended to highlight the idea that whether some reason is a reason to act depends on the circumstances. Although in one circumstance the fact that I am hungry would be a reason to eat dinner, in another circumstances it would not be, for example, if I was about to have surgery and my doctor had instructed me not to eat anything for twenty-four hours before surgery.
[41] Williams, "Internal", 102.

> Now in fact it is initially quite difficult to see why Williams says any of this [referring to a couple quotes of Williams stating his relativistic claim] at all. For, as we have seen, what the internalism requirement suggests is that claims about an agent's reasons are claims about her *hypothetical* desires, no claims about her *actual* desires. The truth of the sentence 'A has a reason to φ' thus does not imply, not even 'very roughly', that A *has* some motive which will be served by his φ-ing; indeed A's *motives* are beside the point…What the internalism requirement implies is rather that A has a reason to φ in certain circumstances C just in case he would desire that he φs in those circumstances if he were fully rational.[42]

In the above we see that Smith sees that internalist thesis as being distinct from the relativism claim. Therefore, one can accept internalism while also rejecting the claim that reasons^{E} are relative to an agent's motivations.

The question to look at now is how Korsgaard and Smith come to their general interpretation of internalism. We will then be able to see why they think the relativism claim is not essential to internalism. (In the quote above we only see *that* Smith does not take it to be essential.)

Korsgaard and Smith's interpretations are shaped largely by Williams's claim that in determining what there is reason for an agent to do we should correct for an agent's beliefs and reasoning. Williams's says that we should do so, because internalism is "concerned with the agent's rationality"[43]. Both Korsgaard and Smith take "rationality" to mean *practical* rationality.

[42] Smith, "Internal Reasons", 117.
[43] Williams, "Internal", 103.

As a result, they take internalism to be concerned with what the agent should do, were the agent fully practically rational.

How we interpret the meaning of Williams's use of "rationality" is pivotal to determining the correct interpretation of internalism. So, I want to point out a possible ambiguity with the notion of rationality being used. When Williams says that we are concerned with an agent's rationality, he could mean that in two different senses. The first sense is *theoretical* rationality. The second is *practical* rationality. To be fully theoretically rational, it is only required that we have all true beliefs and no false beliefs.[44] But an agent that is fully theoretically rational may not be fully practically rational. To be fully practically rational *might* require more than being fully theoretically rational. Whether it does depends on the correct conception of practical reason. Some conceptions of practical reason are "thicker" than others. Instrumentalism, a rather "thin" conception, concludes that an agent who is theoretically rational is thereby necessarily practically rational. To be fully practically rational requires nothing more than to have your desires corrected to account for false beliefs. An Aristotelian conception of practical reason however, allows for an agent being fully theoretically rational without being fully practically rational. An agent might have all true beliefs, and thereby know what there is most reason to do, but whether he is fully practically rational depends upon whether he has the desire to perform that action. If he does not have the desire, then he is not fully practically rational.

Smith and Korsgaard both take Williams to mean fully *practically* rational when he says we are concerned with an agent's rationality.[45] I think that is a mistake. Williams should be understood to mean merely *theoretically* rational (and I also do not think that we are to consider

[44] There might be a concern about whether it is possible for an agent to have *all* true beliefs, so we might modify it to all true beliefs which it is logically possible to have.
[45] Smith, "Internal", 112. Korsgaard, "Skepticism", 15.

what there would be reason for the agent to do, were he *fully* theoretically rational).[46] I will argue for that in the third chapter, but for now my objective is just to draw attention to these two possible interpretations of rationality. Which interpretation is correct significantly affects the nature of internalism. But, with this ambiguity noted, we will now look at Korsgaard and Smith's practical rationality interpretation of internalism.

As I mentioned before, both Korsgaard and Smith accept internalism (as they interpret it). However, they disagree that it entails that all reasons^{E} are relative to an agent's S. To see why they reject the relativity claim, we will consider Smith's explication of William's account of what it is to be fully practically rational. Doing so will allow us to see what Korsgaard and Smith's objections are.

Smith says that on Williams's account of practical rationality, an agent must satisfy three conditions to be fully practically rational:

1. The agent must have no false beliefs
2. The agent must have all relevant true beliefs
3. The agent must deliberate correctly[47]

Condition #2 allows that an agent can be fully practically rational, even if not fully theoretically rational. As long as the lack of true beliefs does not distort the agent's practical conclusions, the agent can still be fully practically rational. For example my lack of true belief about how many

[46]That is because, if we were, the reasons of an agent who lacks information would then be the same as the reasons of an agent who has full information. If an agent lacks information, often there is a reason for them to find out the information. But, if we are to determine what there is reason for the agent to do, were he *fully* theoretically rational, there would not be a reason for the agent to find out the information since he already has it. That is why Williams only claims that the agent is to have all *relevant* information.

[47] Smith, "Internal", 112.

stars there are in the universe does not seem to affect whether I have reason to eat lunch. Therefore, I do not need to have that true belief in order to be fully practically rational. But, if my lunch has been poisoned by an enemy, that information does seem to be relevant to whether there is a reason for me to eat my lunch. So, I would need to know that.

Smith agrees with these three conditions, and it seems clear that Korsgaard does as well. However, they disagree with Williams over the implications of the correct deliberation condition. On their views, Williams underestimates the full potential for rational deliberation to determine what there is reason for an agent to do. That underestimation is responsible for Williams's conclusion that all reasons for action are relative to an agent's S. Korsgaard and Smith give two different explanations as to how Williams's view is deficient.

Korsgaard claims that the cause of Williams's relativism claim is his *content skepticism*.[48] That is, he is skeptical about the ability of practical reason to provide guidance about our actions independent of our desires. Korsgaard thinks that there is such a thing as *pure practical reason*—that is, practical guidance which is determined by reason alone, independent of an agent's desires. Through practical deliberation we can discover what we are rationally required to do, and the conclusions of such deliberation will be the same for every *fully rational* agent (presumably agents who are in the same circumstances). And this is true irrespective of the desires with which they begin. If it is true that pure practical reason can determine what we ought to do irrespective of our desires, it is not the case that all reasons for action are relative to our desires.

Smith makes a similar objection to Williams's conception of correct deliberation. He claims that Williams neglects the role that *systematically justifying* our desires can play in correct deliberation. Here he draws on Rawls's conception of 'reflective equilibrium'. Reflective

[48] Korsgaard, "Skeptisicm", ???.

equilibrium is a state in which our evaluative beliefs are coherent and unified. Someone who believes both that it is *always* morally wrong to lie, and that it would be morally permissible to lie to a would-be-murderer has not achieved reflective equilibrium.

Smith claims that reflective equilibrium amongst our evaluative beliefs is a necessary condition for full practical rationality. However, he also thinks that full practical rationality requires reflective equilibrium amongst our desires. Given that we often find ourselves with conflicting desires, we might try to arrange and/alter our desires into a more coherent and unified set. Once we undergo this process of systematic justification, according to Smith we will (if fully rational) come to have a "maximally coherent and unified desire set" of desires.[49] And, he claims, all fully rational agents (who are in the same circumstances) will have the same set of desires, and so therefore the same reasons to act. Therefore, contra Williams, all reasons for action are not relative to an agent's S.

So, on Korsgaard and Smith's interpretation of internalism, it is only the claim that practical reasons must be capable of motivating fully rational agents. The claim that reasons are relative to an agent's S is distinct from internalism itself, and only follows from theories of practical reason which have a more limited view of the capacity of practical reason to provide guidance for our actions beyond our desires. Given that Korsgaard and Smith think that practical reason has the capacity to determine what there is reason to do apart from our desires, they reject the relativism claim.[50]

The question we must now ask is whether Korsgaard and Smith's interpretation of Williams's internalism is accurate. As I said before, there are strong reasons for thinking that the correct interpretation of internalism upholds both (R) and (N). As should be clear now, however,

[49] Ibid., 117.
[50] Note that those who are externalists on the first interpretation of internalism—that it is the claim that all reasonsE are relative to the motivations of an agent—are likely to find internalism on this interpretation acceptable.

Korsgaard and Smith's interpretation conflicts with (R). They allow that internalism is consistent with the claim that reasons are not relative to an agent's subjective motivational set. We therefore have good reason to reject this interpretation, unless no better interpretation is to be found. If it turns out to be the best interpretation, then Williams just misunderstood the implications of internalism.

There is a third possible interpretation of Williams which also interprets internalism as being concerned with reasons^{E}. It would take both (R) and (N) to be essential claims of internalism. The first interpretation of internalism took (R) to be its essential claim, and concluded that internalism must be presupposing a quasi-instrumental theory of practical reason. It therefore denied that (N) was essential to internalism. The second interpretation took (N) to be essential to internalism, and so concluded that (R) was not. The third interpretation makes both (R) and (N) essential to internalism.

However, on this interpretation, it is almost certain that internalism is just an incoherent theory. There does not seem to be a way to reconcile both (R) and (N) on the reasons^{E} interpretation. If it is going to be true that all reasons^{E} are relative to an agent's S, then it must be that (N) is false. And, if no theory of practical reason is presupposed, then the truth of (R) must be indeterminate. But "indeterminate" is not what internalism seems to be after. As Williams put it, according to internalism, "*by definition*, all reasons are relative".[51] So, if they are incompatible claims, internalism is an incoherent thesis and should be rejected.

However, I think we should consider another option. Instead of interpreting internalism as being concerned with reasons^{E}, we should interpret it as being concerned with reasons^{H}, and in particular, what an agent has most reason to do. Explaining and defending that interpretation is the subject of the next chapter.

[51] Williams, "Internal", 102.

Conclusion

In summary, we have looked at the distinction between practical reasons that there are, reasons that an agent has, and the reasons an agent has most to act upon. I then argued that two claims, (R) and (N), are most likely essential to the internalist position and should be a part of the correct interpretation of internalism, unless no better theory can defended. Next we considered the two primary general interpretations of internalism, neither of which uphold both (R) and (N). I then concluded that we therefore have a strong *pro tanto* reason to reject them. The question now is whether an alternative interpretation of internalism can accommodate both (R) and (N) (as well as the bulk of Williams's writings related to internalism). In the next chapter I will offer an interpretation which I think can meet those demands.

Chapter 2: The Reasons^{H} Interpretation of Internalism

In the last chapter we examined the two primary ways that internalism has been interpreted. Both interpretations take internalism to be a claim about reasons that there are—that is, reasons^{E}. We also saw that neither interpretation is able to accommodate both of the claims which I argued are essential to internalism. Neither is able to uphold both (R), that all reasons to act are relative to an agent's subjective motivational set *and* (N), that no theory of practical reason is presupposed by internalism. In this chapter I will argue that we should interpret internalism to be a claim about reasons^{H}—what an agent *has* reason to do. On the reasons^{H} interpretation, internalism is the claim that an agent's inability to be motivated by a reason^{E} to φ precludes the agent from *having* a reason to φ. To put it another way, the claim is that psychological limitations constrain what reasons an agent has to φ. If an agent lacks a particular element in his S which is necessary for being motivated by a reason^{E} to φ, then that agent has a psychological limitation.[1] One consequence of this interpretation (as we will see later) is that it is consistent with both (R) and (N).

I have two aims in this chapter. The first is to provide a more detailed explanation of what internalism is on the reasons^{H} interpretation. An important aspect of the explanation will be to highlight the impact the reasons^{H} interpretation has for how we are to understand some ambiguous passages of Williams. The second aim of this chapter is to provide three

[1] Although the claim that the agent has a psychological *limitation* may appear to be a negative claim, it does not have to be understood in that manner. Williams, in "Moral Incapacity" in *Making Sense of Humanity*, (Cambridge: Cambridge University Press, 1995), 46-55, considers an agent who is incapable of performing what is (or at least what he perceives to be) an immoral action. Take, for example, someone who is incapable of committing armed robbery merely for the sake of stealing someone's pocket change. In such a case, we may not want to construe the psychological limitation as negative, i.e. as a deficiency of the agent's practical rationality; but instead as a positive, as an indication that the agent *is* practically rational. An ability to commit armed robbery in such a circumstance may be an indication that the agent is not fully practically rational.

considerations in support of the accuracy of the reasons^{H} interpretation. The first consideration is that, unlike the predominant interpretations, the reasons^{H} interpretation *is* consistent with (R) and (N). The second is that the reasons^{H} interpretation provides the most plausible interpretation of Williams's claim that there is an interrelationship between explanatory and normative reasons. And third, I will show that his argument against external reasons (which relies upon the supposed interrelationship between explanatory and normative reasons) is sound on the reasons^{H} interpretation; and, importantly, its soundness does not depend upon any particular theory of practical reason. Therefore, it is a charitable interpretation of Williams's internalism. Assuming these three considerations are established, we will have good reason to think that the reasons^{H} interpretation is an accurate interpretation of Williams's internalism. In Chapter 3 I will provide an argument that with respect to Williams's argument against external reasons, not only is the reasons^{H} interpretation charitable, it is also the *most* charitable interpretation. Those four considerations taken together will constitute a very strong argument that the reasons^{H} interpretation is the correct interpretation. But we first need to take a look at what the reasons^{H} interpretation is.

I. Internalism as a necessary condition for an agent having a reason to φ

According to Williams's last formulation of internalism, "[an agent] *A* has a reason to *φ* only if there is *a sound deliberative route* from *A*'s subjective motivational set…to *A*'s φ-ing"[2] (italics Williams's). People have often thought that his internalism is a sufficient condition account of reasons for action. But that is not the case. In both "Internal Reasons and the Obscurity of Blame" and "Postscript" Williams explicitly mentions the necessary/sufficient

[2] Williams, "Postscript", 91.

condition distinction and asserts that internalism is only the statement of a *necessary* condition of reasons for action.[3]

In this section, I want to provide a detailed explanation of Williams's internalism—as well as this formulation of it—on the reasons^{H} interpretation. On the reasons^{H} interpretation of internalism, internalism is the claim that a psychological inability to be motivated by a reason^{E} to φ nullifies that reason from being one that an agent *has*. To better understand the nature of that claim, we can begin by considering how physical limitations constrain the reasons an agent has. That physical limitations constrain the reasons an agent has is not controversial. As we saw in the first chapter, if an agent has a reason to φ, he must be capable of φ-ing. And so, if an agent does not have the physical capacity to perform an action, he cannot have a reason to perform it. For example, if someone is not strong enough to φ, then he cannot have a reason to φ. There may be very good reasons^{E} to φ, but if he is not strong enough to do so, then he does not have a reason to φ.

Compare the reasons that Superman can have which ordinary humans cannot. Because Superman is physically stronger, faster, and so on, he has fewer physical limitations than ordinary humans. As a result, he can have reason to perform actions which most people do not. If a bus has had its brake lines cut, Superman can have a reason to stop the bus with his own strength—landing in front of it and putting his hand out to stop it. Ordinary humans cannot have such a reason. Depending on the circumstances, an ordinary human might have a reason to try to stop the bus by other means, e.g. by laying down spike-strips, moving a car into its path to slow

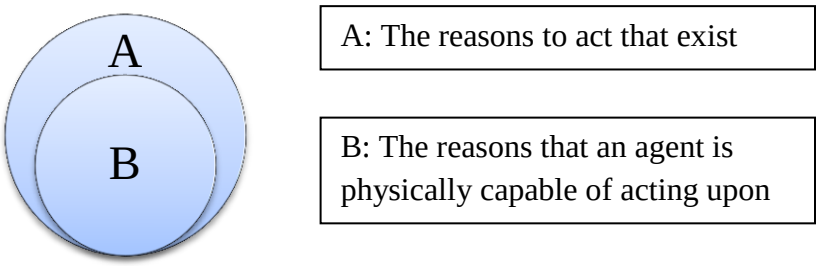
[3] Williams, "Postscript", 92. Williams, "Internal Reasons", 35-6. Notably, in IROB Williams goes on to say that he thinks the *formulation* of internalism is also a sufficient condition, but he asserts that that further claim is *not* a part of internalism. In other words, *Williams* thinks that if there is a sound deliberative route from an agent's motivational set to his φ-ing then that is a normative reason for the agent to perform the action. But he is not defending that claim as part of internalism.

it down, etc. But what he does not have is a reason to step in front of the bus and stop it by holding his arm out.

We can also consider a scenario in which differences in physical limitations between ordinary humans results in a difference in what each of them has reason to do. Two people may be in almost identical circumstances, but because one has physical limitations that the other does not, there can be a significant difference in what each has reason to do. Consider a scenario in which a firefighter and an elderly person are outside a house that is on fire and in which a baby is trapped. The firefighter who is in excellent shape may have reason to storm the house in order to rescue the infant from the fire. But, the frail elderly person who cannot walk and can only get around with the help of a wheelchair does not. Because the elderly person cannot storm the house in order to rescue the infant, he cannot have a reason to do so. He may have reason to do other things to try to save the infant, such as calling the fire department if it is not already there, but he does not have a reason to run into the burning house to rescue the baby. As well, there may even be most reason^{E} to save the infant—that may be what is most important to do in those circumstances—but anyone who does not have the physical capacity to save the infant cannot have a reason to do so.

Below is a diagram which represents the effects of physical limitations on an agent. The whole circle, "A", represents all of the reasons to act in a particular circumstance, i.e., all the reasons^{E}. What generates these reasons is determined by whatever is the correct theory of practical reason. It may be agents' desires, the good that can be promoted, etc. It is likely the case that some reasons^{E} to act in "A" outweigh other reasons, but so long as some consideration is a reason to perform an action, it is a member of "A". But the main thing to understand is that it includes all reasons^{E} for action. The smaller circle, "B", represents the reasons to act that a

particular agent is physically capable of acting upon. Because having a reason requires being capable of acting upon the reason, any reason that is in "A" but is not in "B" is not a reason that the agent has. Consider the elderly person above. There is a reason^{E} to save the infant in the burning house, and so that reason is in "A". But, because the elderly person is physically unable to do so, that reason is not in "B". The elderly person does not have a reason to save the child, because he is physically incapable of doing so.



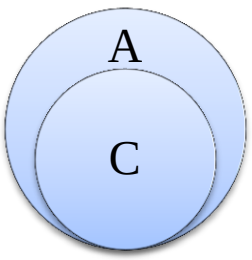
1. The lack of motivation as a limitation on an agent's capacity to act

That physical limitations constrain what we have reason to do is not controversial. In arguing for internalism, Williams is arguing for the apparently controversial (or, more likely, misunderstood) claim that psychological limitations, like physical ones, also constrain what reasons an agent has. Whereas physical limitations are physical incapacities to perform an action, psychological limitations are psychological incapacities to be motivated to perform an action. So, if an agent is incapable of being motivated to φ, then the agent does not have a reason to φ. And, as I pointed out in Chapter 1, even if the agent has some motivational element related to a reason *r*, that is not sufficient for *r* being in "B". If other motivational elements and/or other reasons for

acting prevent the motivation to φ for *r* from being effectual, then *r* is not a reason that the agent has.

Note the senses of "capable" and "possible" and "can" used in the discussion of physical limitations. When we have said that an agent is physically capable of φ, or that it is physically possible for the agent to φ, etc., what we have meant is that it is within the agent's physical capacity to perform the action. The general notion of "possibility" in use here is what it is within an agent's capacity to do. We are not concerned with what is logically, or metaphysically, or rationally possible. These more sophisticated notions of possibility are not in use here. (This will be important later.) The same type of possibility used here will be used in the discussion of psychological limitations and what it is possible for an agent to be motivated to do. So when we say that for an agent to have a reason to act the agent must be capable of being motivated to φ, we mean capable in the ordinary, everyday sense of the word.

For clarity, we can represent the internalist claim with a diagram similar to the one used in representing the effect of physical limitations on the reasons an agent has. In this case it is the psychological limitations of an agent that affect the reasons an agent can have. In this diagram, like the first diagram, "A" represents all of the reasons^{E} to act in a particular circumstance. What generates these reasons, e.g. desires, goodness, etc., is determined by whatever is the correct theory of practical reason. We will represent the reasons^{E} that an agent is psychologically capable of acting upon with "C", to distinguish it from the physical limitations diagram.



A: The reasons to act that exist

C: The reasons that an agent is psychologically capable of acting upon

Since internalism is only a necessary condition, and not a sufficient one, reasons^{E} within "C" are not necessarily also reasons^{H}. They are only *possibly* reasons^{H}. In part that is because there may be other reasons as to why a reason^{E} that is within "C" is not a reason^{H}. For example, it may be that an agent is psychologically capable of acting upon a reason, but not physically capable. To have a reason, an agent must be capable of acting upon it all things considered. So, an agent must be capable both psychologically and physically of performing the action. To be a reason^{H}, not only must a reason^{E} be in "C", it must also exist in "B". A reason that is in "C" but not "B" is not a reason that the agent has. For example, the elderly person considered previously may have the motivation to storm the burning house to save the infant, but because he is not physically capable, he does not have a reason to do so.

This diagram also helps make it more perspicuous that internalism is not a claim about what generates reasons^{E} for action, but is only a claim about what is necessary for an agent to have a reason. So far as internalism is concerned, an account of what generates reasons for action, i.e. what makes it the case that there is a reason^{E} to φ, may or may not depend on an agent's S. There may be numerous reasons^{E} to φ unrelated to an agent's S. But, if a particular agent is incapable of being motivated by any of those reasons^{E}, then none of them are reasons that the agent has for acting. None of them are reasons^{H}. In other words, Williams's internalism is internalism about reasons^{H}, not reasons^{E}. And internalism about reasons^{H} is entirely compatible with a non-instrumentalist conception of reasons^{E}. To put it more simply, internalism about reasons^{H} is consistent with an externalist conception of reasons^{E}.

Although psychological limitations have the same constraining effect as physical ones, identifying an agent's psychological limitations is much more difficult. To get a clearer

understanding of the internalist conception of how to determine what counts as a psychological limitation and what does not, we need to look at two things: 1) the elements that compose the subjective motivational set, and 2) the role of the sound deliberative route qualification (and what it allows us to say about the reasons an agent has). We will take a look at each of these in turn.

2. The elements of the subjective motivational set

To determine what psychological limitations an agent has, we have to determine what can motivate him to act. Those aspects of an agent's psychology that can motivate him to act are the elements which compose his S. Some philosophers have thought that, on Williams's view, S is composed *only* of the *desires* of an agent. However, there is no such restriction for Williams. As long as something can play a role in an agent's motivation to act, then it can be and is an element of the agent's S. As Williams has stated, the agent's S is composed of "desires…dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects…embodying commitments of the agent."[4] It is not just desires in that are included in S. Think of S as composed of *whatever* it is that plays a psychological role in motivating an agent to act. To put it another way, whatever it is that affects a particular agent's capacity to be motivated by a reason is to be included in the agent's S. If the agent is motivated by pleasure to act, then that is a part of his S. If the agent is motivated by moral considerations, then that is a part of his S. If the agent is motivated by opportunities to create pieces of art, then that is a part of his S. And, importantly, these motivations do not have to be desires. It does not have to be that the agent has a desire to act morally, for example. The agent may have no preexisting desire to act morally, but instead

[4] Williams, "Internal", 105.

may have a psychology such that upon recognition of a moral reason, he will come to be motivated to perform the moral action. As Williams states in IROB, "There is, then, no attempt to exclude altruistic or other ethical considerations from the rational[5] agent's *S*. For most agents, those patterns of motivation appear together with many others—desires, projects, sympathies and so on."[6] So, if an agent is concerned with moral considerations and can be motivated to act in accordance with them, then that is a part of the agent's S.

My claim that we should think of S as being composed of whatever is capable of motivating an agent—even pure reason—is likely to strike some as an inaccurate representation of Williams's position. After all, although Williams does provide a list of motivational elements that can be found in an agent's S in addition to desires, the various elements he does mention, e.g. projects, commitments, dispositions of evaluations, etc., could all plausibly be construed as types of desire. However, in Williams's text I think there are the resources for an argument which shows that he does allow for elements which are not desires.

According to Williams, internalism allows for the possibility of, as Christine Korsgaard puts it, "pure practical reason". Pure practical reason, which gets traced back to Kant, is the idea that "the structure and not simply the content of practical reason can ground reasons."[7] If there is such a thing as pure practical reason, Williams claim that "then it would be true of every rational agent that there was a sound deliberative route from his or her *S* to actions required by such reasons."[8] By Williams allowing that there could be such reasons, and then claiming that there would then be a sound deliberative route from any rational agent's S to actions required by pure

[5] Later I will argue that Williams's use of "rational" here is misleading. He is not concerned with what an agent's S would look like if the agent were fully rational. Instead, he means what an agent would be motivated by when his S is purged of any desires that rest upon a false belief. There is a substantial difference between these two notions.
[6] Williams, "Postscript," 92.
[7] Ibid., 94.
[8] Ibid.

practical reason, it seems he must say that the capacity to be motivated by pure practical reason is a motivational element in the agent's S. If not, then he would be committed to the idea that an agent could be motivated to perform an action for which he does not have a corresponding element in his S. But that would constitute an acceptance of externalism. (And, remember that he claims internalism is consistent with the existence of pure practical reason.) In "Internal Reasons and the Obscurity of Blame", Williams describes externalism as the claim that an agent can have a reason to φ "without there being any shadow or trace of that [motivational element] presently in his S."[9] Also, importantly, if what is responsible for the element's existence in the agent's S is the fact that the agent is capable of being motivated by pure practical reason, then it would not be accurate to describe that element as a desire (narrowly construed). It falls under the much broader category of a motivational element—in that it (the capacity to be motivated by pure practical reason) is capable of motivating the agent to act. So, given that Williams claims pure practical reason is compatible with internalism, he must allow for non-desires to be elements in an agent's motivational set.

Although the internalist view is very liberal about what elements can be in the motivational set, it does require that the elements actually be in the agent's S. We cannot say that, because a fully rational agent would be motivated to act morally, that therefore moral motivations are in each agent's S. So, if an agent does not care about morality and would not be motivated by an action that is the moral thing to do, then moral considerations are not a part of the agent's S.[10] And this is so, even if a concern for moral considerations is necessarily part of

[9] Williams, "Internal Reasons", 39.

[10] Whether an agent can be morally required to perform an action even if he is incapable of being motivated by it is a controversial issue. However, on the internalist view, the agent does not *have* a reason, because the agent is not actually capable of acting upon the reason, i.e. the agent cannot act upon the reason. Notice that the internalist conception of "can" here is different from the compatibilist conception of can in "ought-implies-can". This will be discussed in much more detail in Chapter 4.

any *fully practically rational* agent's S. As Williams writes in "Postscript", it may be that an agent is incapable of acting upon a reason because he does not have a related element in his S, and it may be that the agent is *defective* for not having that element in his S. But, Williams adds, since the agent does not have the related element, the reason is not one that he has.[11] What elements compose an agent's S depends on the actual agent's S, not the ones that would exist in a fully practically rational agent's S.

3. The purposes of the sound deliberative route

The inclusion of the sound deliberative route in the formulation of internalism serves two purposes. The first is to filter out apparent motivations from inclusion in the agent's S when those motivations depend upon a false belief, or the lack of a true one. The second purpose is to allow that an agent can satisfy the internalist requirement for having a reason to φ, even if he does not presently have the particular motivations necessary for φ-ing. We will take a look at each purpose in turn.

Although what motivational elements compose a particular agent's S is an empirical issue, internalism does not take (apparent) motivations at face value. Sometimes agents are motivated to perform an action only because either they have a false belief or lack a true belief relevant to the action. So, the first purpose of the sound deliberative route is to ensure that when we are determining what an agent's actual psychological limitations are, we are determining

[11] Williams, "Postscript", 96. A fuller (though indirect) explanation and defense of this claim can be found in Chapter 3, II.3, starting on p. 122. Although the section is directly related to the issue of whether Williams claims that all reasons—both reasons^{E} and reasons^{H}—are internal, it can also be seen from the discussion that Williams, in determining the elements of an agent's motivational set, is concerned with the motivations of the actual agent, not a fully practically rational version of the agent.

them based on the agent's actual motivations, and not the motivations that he merely appears to have as a result of false beliefs (or the lack of a relevant true belief).

Consider Williams's gin and tonic example. The agent in question wants to take what is in the glass and mix it with tonic and drink it. But he wants to do that only because he (falsely) believes that gin is in the glass. However, what is in the glass is petrol. So, does the agent have a motivation to drink what is in the glass, i.e. to φ? According to Williams, no. The agent does not *really* want to drink the petrol that is in the glass. His motivation is for a gin and tonic, and he only *thinks* that gin is in the glass. Were the agent to deliberate about whether to drink what is in the glass on the basis of true information, i.e. were he to be deliberating soundly, he would not be motivated to drink it. Therefore, this agent does not meet the internalist requirement for having a reason to drink what is in the glass. (There is also a more obvious reason why he does not, and that is because to have a reason, there must also be a reason^{E}. And, there is most likely not a reason^{E} to drink petrol mixed with tonic.)

We can also consider a scenario where an agent appears to lack the motivation to φ, but actually has it. A high school student may not want to go to a high school dance because he believes that the girl he likes is not going to be there. However, she actually is. At present he does not appear to be motivated to φ, but that is only because he has a false belief. Were he to know that she was going to be there, he would be motivated to go to the dance. In this case there is a sound deliberative route from the agent's motivations to his φ-ing. That is, given the agent's motivational profile, there are truths which, when combined with it, would lead to the agent's φ-ing. So, the apparent lack of motivation to φ is not really a lack of *motivation*, but instead it is the *lack of a true belief*. The agent in this case has the motivation; he just does not have the true

belief necessary for performing the action. Since the agent has the motivation, he satisfies the internalist requirement.

When Williams says that we can correct for an agent's beliefs because we are concerned with the agent's rationality, some philosophers have thought that we are concerned with identifying what the agent would be motivated to do, were the agent fully practically rational.[12] Were that the concern of internalism, the objections of Shafer-Landau et al., as well as Korsgaard and Smith would be relevant. But, if I am correct, the correction of an agent's beliefs is only intended to determine what the agent's actual motivations are—not what they would be if the agent were fully rational. The point of doing so is to determine what the agent's actual psychological limitations are. Once we determine that, we can determine whether he meets the internalism requirement for having a reason (namely, having the motivation to act on it).

However, just because an agent meets the internalism requirement for having a reason to φ, that does not entail that the agent actually has a reason to φ. We have to remember that the internalist requirement is only a necessary condition for having a reason; it is not a sufficient condition. Having a reason to φ requires more than just meeting the internalism requirement. It requires that acting upon the reason^{E} to φ be within all of the agent's capacities. Another requirement, as we have already seen, is that the agent must be physically capable of φ-ing. We might call this the *physical requirement*.

An important implication of the fact that internalism is only a necessary condition is that, in order for an agent to actually have a reason to φ, not only must there be a sound deliberative route from the agent's S to his φ-ing for that reason, but he must also be capable of following that deliberative route. We might call this the *intellectual requirement*. So, if an agent cannot become aware of the reason to φ, then it is not within his capacity to φ for that reason. For

[12] In Chapter 3 I will explain why that is an erroneous interpretation of Williams.

example, if an agent has a false belief which prevents him from recognizing the reason, (and it is not within the agent's capacity to find out the truth), then the agent does not have a reason to φ. In the dance scenario above, due to the agent's false belief that the girl will not be at the dance, he is not capable of following the deliberative route that would lead to his φ-ing, and so he does not have a reason to go to the dance. Or, if the agent has all of the true information he needs, but does not have the intellectual ability to deliberate on the basis of the information to reach the conclusion to φ, then he also would not have a reason to φ.

That Williams requires the agent to be capable of following the route in order to actually have a reason is often overlooked. But, we see this in two places. First, in "Internal and External Reasons" he writes that although an agent can at times be unaware of a reason he has for acting, "[f]or it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that *A* would have a reason to *φ* if he knew the fact."[13] Unfortunately it is not clear what Williams means by the unknown fact being "fairly close and immediate".[14] But, for our present purposes what is important is that we see that an agent's lack of information can affect what he has reason to do. Presumably, that lack of information would prevent him from being able to follow the sound deliberative route, and so he does not have a reason to act on it. We also see Williams's requirement that an agent be capable of performing the deliberation in order to have a reason to φ in his *Ethics and the Limits of Philosophy*. Williams writes that an agent who (at least seemed to) have an obligation to perform an action may not have *had a reason* to perform it. He

[13] Williams, "Internal", 103.
[14] I suspect that by the fact being "close and immediate", he means that it is a fact which the agent has the capacity to come to be aware of.

then goes on to give various ways in which he could fail to have a reason. One of them is a "general deliberative incapacity".[15]

I suspect that there might be a concern that the case for my claim that Williams thinks that having a reason requires an agent to be capable of following the deliberative route rests on too little textual support. I agree that the textual support is minimal, but let me add a further consideration in its favor. I think the lack of support is due to the fact that—if my interpretation is correct—in his articles Williams is primarily concerned with explaining and defending the internalist requirement for having a reason. And, with respect to it, there is not a requirement that the agent be capable of following the route. So, it is not that surprising that he did not mention it more.

Given the complexity of the present issue about whether an agent must be capable of following the sound deliberative route, it will probably help to reiterate the point being made. With respect to the *internalism requirement* for having a reason, it is *not* necessary for the agent to be able to follow the sound deliberative route from his S to φ-ing for the reason in question. If there is a sound deliberative route, that indicates that the agent has the *motivational capacity* necessary to φ for the reason in question. And the internalism requirement is only the requirement that an agent have the motivation necessary to perform an action. But, in order for the agent to actually have a reason, he must be capable of performing the action all things considered, and so, if he is intellectually incapable of following the route, then he does not have a reason to φ for that reason. In the latter case he would meet the internalism requirement for having a reason to φ, but not the intellectual requirement.

The second purpose of the sound deliberative route is to allow that agents can meet the internalist requirement of being capable of being motivated to perform an action even if they do

[15] Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge: Harvard University Press, 1985), 192.

not presently have the *particular* motivation(s) necessary for performing the action. Having a particular capacity (e.g. physical, intellectual, motivational, etc.) to perform an action does not require having all of the capacity *now* (unless the action in question is to be performed at the present time). That is because it may be possible for the agent to acquire the capacity to perform the action.

Take physical capacity as an example. When we say that an agent must be physically capable of performing an action, we mean that it must be physically possible, given what it is within his power to bring about. And it may be within an agent's physical capacity to φ even if the agent does not at present possess the physical capacity necessary for φ-ing. For example, an agent may have the physical ability to change a flat tire on a vehicle, even if the agent is not capable of lifting the vehicle using his own power. If there is a car jack in the trunk which the agent can get, the jack will provide the additional "strength" that he needs to lift the car. But, if there is no car jack in the trunk, and there are no other resources for lifting the car which are within the agent's ability to find, then the agent does not have a reason to change the tire. And this is true, even if, unbeknownst to the agent, in a few minutes some Good Samaritans will decide to stop by who have a car jack. It is not within his capacity to bring that about, and so it is not within his capacity to lift the car (prior to their arrival). So, before they arrive, he does not have the physical capacity. Afterward, he does.

This line of thinking also applies to an agent's motivational capacity. Even if the agent does not presently have all of the motivations necessary for performing an action, if it is within his capacity to acquire the necessary motivations, then he meets the internalism requirement. Consider the following scenario. Suppose that Smith and Jones have a broken relationship, and it was due to a trivial disagreement. If they have had a close and extended relationship, there is

probably a reason for them to reconcile. However, Smith may be so angry with Jones that he could not at present talk with Jones in order to mend their friendship. So, despite there being a reason to φ (to reconcile), Smith does not have the motivation to do so. Does that entail that he does not have a reason to reconcile? Not necessarily. If Smith has the motivation necessary to take steps towards having the motivation to reconcile, then he could have a reason to do so. For example, he may decide to take some personal time to reflect on the disagreement, knowing that it is trivial and so thinking about it will dissolve much of his anger. If doing that would then generate the motivation to reconcile, then, since it is within his motivational capacity to reconcile, even though indirectly, he can at present have a reason to reconcile in the near future. (But, if the issue is whether he has a reason to reconcile *now*—that is, that at this very moment he should talk with Jones—then he does not have a reason to do so because he does not now have the motivation to reconcile. It is beyond his motivational capacity to φ now.)

In the example above, it is within Smith's capacity to be motivated to reconcile with Jones, despite not presently having the specific motivation to do so. And, the reason is that there is a sound deliberative route from his current motivations to the motivation to reconcile. By reflecting on his relationship with Jones, and thinking about the trivial nature of the disagreement, Smith is able to acquire the motivation to reconcile with Jones. The inclusion of the sound deliberative route in the formulation of internalism is intended to capture such scenarios. To have a reason to φ for some reason does not require that you have the motivation presently, but instead only that it be within your own capacity to acquire that motivation. And, importantly, this is why Williams says that, "unless a claim to the effect that an agent has a reason to φ can go beyond what the agent is already motivated to do – that is go beyond his

already being motivated to φ – then certainly the term [internalism] will have too narrow a definition."[16]

So, what we have just seen is that internalism—in determining whether an agent has the motivational capacity to act on a reason—is concerned with the actual motivations of an agent, not the ones which he appears to have as a result of false beliefs. And, as we saw, this is not the same as what motivations the agent would have if he were fully practically rational. If an agent does not have the same motivations that a fully practically rational agent would have, then, even if we correct the agent's beliefs, he may not come to have the same motivations which the fully practically rational agent would have. So there is no guarantee—and in fact it seems very unlikely—that once we correct agents' beliefs, they will all come to have the same motivations. Their motivational sets may be different, perhaps significantly. (We might deny this, however, if we hold to a Socrates-like claim that no one knowingly does what is contrary to practical reason.) Because of the variance in their motivations, i.e. their psychologies, the reasons^{E} one person can be motivated by will be different from the reasons^{E} another agent can be motivated by. And, therefore the reasons each agent *has* would be different. That is the basis for Williams's claim that there is "a relativity of the reason statement to the agent's *subjective motivational set".*[17] In other words, whether "A has a reason to φ" is true is relative to an agent's motivational capacities.

Hopefully it is clear at this point that internalism does not make any pronouncements about whether any type of S is ideal or not; nor does it claim that any S is just as fully rational as any other. For all that internalism says it may be that one type of S is the ideal type, whether because it is the most rational, or moral, etc. Whether that is so must be determined by

[16] Williams,"Internal Reasons", 36.
[17] Williams, "Internal", 102.

arguments for the correct account of practical reason. But that is not the concern of internalism. Internalism is concerned with what reasons an actual agent can be motivated by, not with what reasons an ideal agent can be. And so it may be that a genuine reason^{E} is not one that an agent has, precisely because he is not an ideally practically rational agent, and so does not have the ideal S.[18]

It may be helpful to reiterate that the formulation of internalism only states a necessary condition for having a reason to φ. It does *not* state 1) a sufficient condition for having a reason to φ, 2) a necessary condition for the existence of a reason^{E} to φ, or 3) a sufficient condition for the existence of a reason^{E} to φ. In the discussion of the effect that the elements of an agent's S—in particular in the *lack* of elements in an agent's S—has on an agent's reasons^{H}, it is perhaps tempting to think that internalism is claiming that the elements in S generate reasons^{E} or reasons^{H} to act. However, it does not. It only claims that a psychological (i.e. motivational) limitation *precludes* the existence of a reason to act. Just as claiming that physical limitations preclude an agent having a reason to φ does not entail that a physical capacity to φ generates a reason^{E} to φ, the internalist claim that a psychological limitation precludes an agent having a reason to φ also does not entail that a psychological capacity to φ generates a reason^{E} to φ. Internalism only claims that in order for an agent to have a reason to φ, the agent must have some motivation related to the reason in question.

It is not a sufficient or necessary condition for the existence of a reason^{E} because it does not claim what *generates* a reason to act. It does not say that an agent's motivations create a reason^{E} to perform an action, nor that the agent must have a motivation related to a putative reason^{E} in order for it to be a genuine reason^{E} to act. Whether there is a reason^{E} depends on the correct account of practical reason, not internalism. Internalism is not an account of practical

[18] See Williams's explanation of this in "Postscript", p. 96.

reason. It is only a partial account of what is required for an agent to *have* a reason—namely, that an agent has the motivational capacity to act upon it. And because it is only a partial account of what is required for an agent to have a reason, it does not state a sufficient condition for having a reason to φ, only a necessary one. That is because even if an agent has the motivation to act on a reason to φ, he may not have the physical or intellectual capacities to act on it. To have a reason to φ requires that it be within an agent's capacity to φ, and so if the agent lacks a physical or intellectual capacity necessary to do so, then the agent does not have a reason to act on it—despite having the motivational capacity to φ. So, satisfying the internalist requirement does not entail that the agent has a reason; satisfying it only determines that it is *possible* for the agent to have a reason to φ.

In summary, internalism is a necessary condition on an agent having a reason to φ. To have a reason to φ, it must be within the agent's psychological capacity to be motivated to φ for the reason^{E} in question. But, internalism does not include motivations that are dependent upon a false belief as genuine motivations of the agent. And, since it is possible for an agent to deliberate from his current motivations such that he acquires new motivations, the present lack of the particular motivation to φ does not necessarily preclude an agent from having the motivational capacity to φ. For those reasons, the formulation of internalism states that an agent has a reason to φ only if there is a sound deliberative route from the agent's S to the agent's φ-ing for that reason.

4. The nature of the externalist position

If my interpretation is correct, what then is the nature of the externalist position? Who is it that Williams sees himself as arguing against? On my interpretation of internalism, the externalist counterpart holds that psychological limitations do not constrain the reasons that an agent has. That is, a psychological inability to be motivated by a reason *r* to φ does not preclude *r* from being one that an agent has. So, although (by definition) the reasons that an agent has are constrained by his limitations, the externalist denies that psychological limitations constrain the reasons an agent has. For some reason the externalist thinks that psychological limitations do not have the same constraining effect as other limitations of an agent. Given the requirements for having a reason, the externalist position seems contradictory. If the reasons an agent *has* are constrained by what reasons he is capable of acting upon, then it is contradictory to say that the agent has a reason to act upon a reason which he is psychologically incapable of doing. The burden is heavily upon the externalist (on this interpretation of internalism) to explain why psychological limitations do not constrain the reasons that agents have for acting.

That the externalist position is apparently, if not actually, self-contradictory might seem at first to be a reason to doubt the accuracy of my interpretation. We might question why Williams would spend so much time defending a position whose rival affirms such an obvious self-contradiction. This concern gains additional weight when we consider the number of philosophers who have rejected and argued against internalism. How could so many of them accept a self-contradictory position?

There are two responses to these concerns. First, that we might have this worry about my interpretation actually provides some support for it. In IER, Williams claims that the concept of external reason statements are "false, or *incoherent*, or really something else misleadingly

expressed"[19] (italics mine). That externalism as defined by my interpretation is incoherent is an indication, to some extent, that I have interpreted Williams correctly. *Ceteris paribus*, an interpretation of Williams that did not capture those aspects of the externalist position would be less plausible.

Second, if my interpretation is correct, we are able to make sense of the number of philosophers who have rejected internalism. On my interpretation, they have not, or at least very few have, accepted the externalist position which Williams has in mind. As we saw in the first chapter, most philosophers have interpreted his internalism as being a claim about reasons^{E} and not reasons^{H}. Those who have rejected internalism have done so because they misunderstood the true nature of Williams's internalism, not because they have an inability to recognize a self-contradiction. Few of them, if any at all, hold the position that psychological limitations do not constrain the reasons that an agent has to act upon. Rather, those who view themselves as externalists hold that psychological limitations do not constrain the reasons that exist, reasons^{E}.

II. Three considerations in favor of the reasons^{H} interpretation

Up to now, I have largely been concerned with merely explaining my interpretation of Williams's internalism. I will now turn to the task of defending the accuracy of my interpretation. As I mentioned previously, in this chapter I will provide three considerations which support the accuracy of the reasons^{H} interpretation. We now turn to the first consideration.

1. Only the reasons^{H} interpretation is consistent with (R) and (N)

[19] Williams, "Internal," 111.

Now that we have an understanding of the reasons^{H} interpretation, we can examine whether it is consistent with Williams's claims that (R), all reasons for action are relative to an agent's subjective motivational set, and (N), no particular conception of practical reason is presupposed by internalism. Since the reasons^{H} interpretation of internalism is concerned with reasons^{H}, and *not* reasons^{E}, it is consistent with both (R) and (N). Internalism^{H} allows that any theory of practical reason can be correct. That is, it allows that any type of consideration can count in favor of performing an action and therefore generate a *reason^{E}*. Most importantly, it does *not* require that an agent have a motivation related to a consideration for it to be a reason^{E}. It allows that objectivist (i.e. value-based), Kantian, and (not surprisingly) instrumentalist accounts of practical reason can be correct. Therefore it is consistent with (N). And, although it allows for any consideration to constitute a reason^{E} to act, it claims that an agent *has* a reason to φ only if it is within his capacity to be motivated to φ. Therefore, the reasons the agent *has* are relative to his motivational set. And so internalism^{H} is consistent with (R). So internalism^{H} is consistent with (R) and (N). In Chapter 1 I already showed that none of the predominant interpretations are consistent with both (R) and (N). Hence, internalism^{H} is the only interpretation which is consistent with both (R) and (N). That is the first consideration in favor of its accuracy.

2. The interrelationship principle is most plausible on the reasons^{H} interpretation

Williams argues that there is an interrelationship between explanatory and normative reasons. He also claims that that interrelationship provides a fundamental motivation for accepting internalism. The rest of this chapter will be concerned with those two claims. In this

section we will be concerned only with the first claim. To begin, I will examine Williams's various claims that there is an interrelationship between explanatory and normative reasons, which he says is the basis for the internalist position. I will then formulate a succinct statement of that relationship—what I will call the *interrelationship principle* (IP). I will then argue that the interrelationship principle is most plausible on the reasons^{H} interpretation. In fact it is rather implausible on the reasons^{E} interpretation. That will provide some support for preferring the reasons^{H} interpretation over a reasons^{E} interpretation.

In the next section, since Williams claims that the interrelationship principle is a fundamental motivation for internalism, I will then use IP to serve as the foundation for Williams's argument against the possibility of external reasons. The argument will be formulated at a level general enough so as to allow for it to be consistent with the various interpretations of internalism (although slight changes will need to be made, depending on the interpretation). I will then show that on the reasons^{H} interpretation the argument against external reasons is sound, and without presupposing a theory of practical reason. The upshot will be that there are very strong reasons for believing that the reasons^{H} interpretation of Williams is the correct interpretation.

The interrelationship principle. Before I put Williams's repeated claim there is an interrelationship between explanatory and normative reasons into the form of a succinct principle, I first want to examine his initial statement of the interrelationship in IER, some problems with that statement, and finally his revised statement of it in IROB. In IER Williams states the following about the interrelationship between explanatory and normative reasons.

> [The] explanatory dimension is very important, and we shall come back to it more than once. If there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action (it does not follow that they must figure in all correct explanations of their action)."[20]

Williams's claim that people have to sometimes act for a reason in order for it to be a reason has been a source of dispute. Why is it that people *sometimes* have to act for those reasons? Does someone doing it once for that reason fulfill that requirement? Also, is it a requirement of the reason for action for a particular person that *he* sometimes act on that reason, or only that other people sometimes act on that reason? If it were the latter, the idea might be that since some people are capable of acting upon it, if the person who we are considering is not motivated by it, it may just be a deficiency on their part. But why is what is counted as a deficiency dependent upon empirical data as to whether some people actually act on the reason? Perhaps all actually existing persons suffer from the same deficiency. Why can we not determine what counts as a reason conceptually, without having to determine whether actual people act on that reason? Given the problems with this expression of Williams's position, it will prove helpful to rely on his later expression of it in IROB.

In IROB Williams restates the claim that there is an interrelationship between normative and explanatory reasons as the following. He writes:

> It must be a mistake simply to separate explanatory and normative reasons. If it is true that he has a reason to φ, then it must be possible that he should φ for that reason; and if

[20] Williams, "Internal," p. 102.

> he does act for that reason, then that reason will be the explanation of his acting. So the claim that he has a reason to φ – that is, the normative statement 'He has a reason to φ' – introduces the possibility of that reason being an explanation; namely, if the agent accepts that claim (more precisely, if he accepts that he has more reason to φ than to do anything else). This is a basic connection. When the reason is an explanation of his action, then of course it will be, in some form, in his S, because certainly – and nobody denies this – what he actually does has to be explained by his S.[21]

This formulation seems clearer than the first. For one, it has the advantage of only referring to the agent himself, and the possibility of his acting on the reason, not other people. There is not a worry about whether an agent's reasons for action depend upon the possibility of others acting on the reason. Second, it seems to be clearer in explaining that the reason must be a reason that the agent is capable of being motivated by in the circumstance he is in. Whether the agent is capable of being motivated by it in another circumstance is, by omission, deemed irrelevant.

Below I have reduced Williams's claim about the interrelationship between explanatory and normative reasons to two sentences. In the quote above, he uses the locution "has a reason", which might seem to favor my position. However, as I pointed out earlier, he has not been clear as to the specific meaning of that phrase. But, since we know that he is concerned with "normative reasons" we can use that phrase, intentionally leaving it ambiguous as to whether it is reason^{E} or reason^{H}. (One of the objectives of Chapters 2 and 3 is to provide several arguments which provide us with good reason to ultimately interpret it as reason^{H}.) The interrelationship principle is then:

[21] Williams, "Internal Reasons", 39.

IP If there is a normative reason *r* for A to φ, then it is possible for A to φ for reason *r*. If it is possible for A to φ for reason *r*, then it is possible for *r* to be part of an explanation of A's φ-ing.

How we understand this principle depends upon how we interpret "normative" as well as "possible". For now I want to focus on the two possible interpretations of "normative". By doing so, it can be seen that, at least *prima facie*, there is good reason to think that IP is making a claim about reasons^{H} and not reasons^{E}. However, as I will show, my argument is not definitive because there are alternative ways of interpreting "possible" which are also plausible, at least initially. In Chapter 3 I will show why the reasons^{H} interpretation is to be preferred over those alternatives.

Normativity: agent-neutral and agent-relative. "Normative reason" can be interpreted in one of two ways. The first way is *agent-neutral* normativity; or, if you prefer, a norm related to an activity. The second is *agent-relative* normativity.

Agent-neutral normativity is a standard which must be met if an activity is going to be done correctly. The norm is not affected by the capacity of an agent to perform the action correctly. This type of normativity attaches to all types of actions—multiplying numbers, baking a cake, reading a book, and so on. For example, with multiplication, the norms of math require that if one is going to multiply five and seven correctly, one must get the answer thirty-five. This norm is not affected in any way by the capacity of the agent who is trying to perform the multiplication. Even if the agent is incapable of getting the correct answer, the norm still exists. Likewise, with baking a cake or reading a book, even though there might be a bit more leeway in how one might do these and still do them correctly, at some point, were the agent to deviate

enough, he would cease to be baking a cake or reading a book--or at least not doing them correctly.

Agent-neutral normativity might also be applied to the "activity" of performing an action. Like multiplying or baking a cake, the "activity" of performing an action might have certain requirements which must be met in order to be done correctly. We can be concerned with identifying what characteristics or conditions makes an action correct (some might also characterize it as a rational action). Theories of practical reason are often concerned with this issue, and most of them allow that agents can fail to act in accordance with the agent-neutral norms of practical reason. For example, instrumentalists allow that an agent could fail to recognize what means are actually suitable for obtaining his desired end. More robust theories of practical reason allow that an agent can lack a desire necessary for performing the action which there is reason^{E} to perform.

If we are concerned with agent-neutral normativity, we are concerned with *reasons*^{E}. If someone is multiplying "5 x 7" we can say that there is a reason^{E} to get "35". The reason is that it is the correct answer. If someone is baking a cake, we can say that there is a reason^{E} to add eggs to the mixture, a reason^{E} to add flour, another reason^{E} to turn on the oven, and so on. These reasons all exist if the activity is baking a cake. And this is so irrespective of whether the agent trying to bake the cake is capable of acting upon them. And so if there are agent-neutral norms for the activity of performing an action, there can be reasons^{E} to perform an action, even if the agent is not capable of acting upon those reasons^{E}.

The second type of normativity is *agent-relative* normativity; or, if you prefer, a norm related to a particular agent. With agent-relative normativity, we are concerned with what the agent is actually capable of doing, given his capacities and limitations. If an agent is incapable of

performing an action, then there is not an agent-relative normative requirement to perform the action. Consider a five-year old who had seen his mother bake a cake and decided that he wanted to bake one too—all by himself. Although the practice of baking a cake requires that certain steps be followed, in this case it is beyond the capacity of the five-year old to follow them. That is, although there is a reason^{E} to measure out a certain amount of flour and sugar, a reason^{E} to preheat the oven to a certain temperature, and so on, the child is not capable of acting upon those reasons. Therefore we would say that there is not an agent-relative norm requiring that the child actually follow the steps in the cake recipe.

There is also a notion of agent-relative normativity related to the activity of performing an action. We can be concerned with what reasons^{E} a particular agent is capable of acting upon. If an agent is incapable of acting upon a reason^{E} then there is not an agent-relative norm requiring that the agent perform the action. There may be agent-neutral norms for performing the action, but there is not an agent-relative norm.

If we are concerned with agent-relative normativity, we are concerned with reasons^{H}. Reasons^{H}, like agent-relative norms, are constrained by an agent's limitations. If an agent is incapable of acting on a reason^{E}, then that reason is not a reason^{H} for that agent. So, for the child who wants to bake a cake, there is a reason^{E} for him to measure out a certain amount of flour, and a reason^{E} to preheat the oven to a certain temperature, and so on. But those reasons are not ones that the child *has*. This also goes for reasons for action. There may be a reason^{E} for an agent to perform an action, but if the agent is not capable of performing it, then that reason is not one that the agent has.

One possible misunderstanding of the notion of agent-relative normativity is that it merely means what it is *fair* (or something similar) to expect of an agent. There are some actions

which it is within an agent's capacity to do, but it would be too demanding to expect the agent to perform the action. Therefore, we might decide not to place a normative requirement upon the agent to do so. However, that is not the meaning of agent-relative normativity here. Rather, it is concerned with what there is reason for an agent to do, once we have taken the agent's limitations into account. An agent could have reason to perform an action, even if it would be extremely burdensome to do so, since burdensome actions can still be within an agent's capacity.

The notion of normativity in the interrelationship principle. What we now need to determine is which type of normativity Williams has in mind when he says that there is an interrelationship between explanatory and normative reasons. My contention is that it is the agent-relative conception of normativity. That is (at least initially) the most plausible interpretation. To see that it is, we can first consider the implications of the agent-neutral interpretation and see that it is itself fairly implausible. The agent-neutral interpretation of IP I will call IPAN (Interrelationship Principle-Agent-Neutral).

IPAN If there is a reason^{E} *r* for A to φ, then it is possible for A to φ for reason^{E} *r*. If it is possible for A to φ for reason^{E} *r*, then it is possible for *r* to be part of an explanation of A's φ-ing.

If we take this interpretation a very implausible implication seems to follow from it. On this interpretation, *any* incapacity (not just merely motivational) to act on a putative reason^{E} precludes its existence. For example, if an agent is physically incapable of φ-ing, then it is not possible for him to φ for reason *r*. But it is almost undeniable that there can be a reason^{E} for an agent to φ, even if the agent is incapable of performing the action. Even if an agent cannot save

his wife from drowning because he is physically incapable of swimming, there is still a reason^{E} to save her. Since IP denies that claim if interpreted as being about reasons^{E}, we should not interpret it as being about reasons^{E} (though I will explain later why we cannot make this claim conclusively).

Instead, we should take IP to be interpreted to be a claim about reasons^{H}. On that interpretation, we get IPAR (Interrelationship Principle-Agent Relative).

IPAR If A *has a reason r* to φ, then it is possible for A to φ for reason *r*. If it is possible for A to φ for reason *r*, then it is possible for *r* to be part of an explanation of A's φ-ing.

This is a much more plausible claim. Given that our concept of what it is to have a reason to φ requires that the agent be capable of φ-ing, a principle which asserts that it must be possible for an agent to φ is a very plausible claim, if not even trivial. Given that IPAR is intrinsically very plausible, and IPAN is not, the IPAR interpretation of IP is a more charitable one of Williams.

Why then have so many philosophers overlooked the IPAR interpretation? I think it is because of Williams's allowance for us to correct the beliefs of the agent in determining what the agent has normative reason to do. This at least seems to indicate that we are concerned not with what an agent is capable of doing as he currently is (which is relevant to determining reasons^{H}), but instead what an *idealized version* of the agent is capable of doing (in particular, what the idealized agent is capable of being motivated to do).

This allowance for the correction of the beliefs of an agent suggests that perhaps a different interpretation of "possibility" is in use. On the initial interpretation, "possibility" meant "within the agent's capacity". But, given the allowance for the correction of the agent's beliefs,

many have taken "possibility" to mean "rational possibility"—that is, what the agent would do were he fully rational.[22] Philosophers who accept one of the two predominant interpretations of internalism largely accept this understanding of "possibility".[23]

Since there is an alternative conception of "possibility" that might be in use in IP, my argument that IP is more plausible when interpreted as IPAR than IPAN loses some of its strength. IPAN would now read:

IPAN^{RP} If there is a reason^{E} *r* for A to φ, then it is *rationally* possible for A to φ for reason^{E} *r*. If it is *rationally* possible for A to φ for reason^{E} *r*, then it is *rationally* possible for *r* to be part of an explanation of A's φ-ing.

This reading only requires of reasons^{E} that fully rational agents would act on them. It does not require that less-than-fully rational agents would. Therefore, the limitations of a less than fully rational agent do not constrain what reasons^{E} there are. This does make the interpretation of IP as being about reasons^{E} more plausible. Consider its implications for the math example. That the less-than-fully-rational first-grader is not able to get the answer "35" when multiplying "5 x 7" does not entail that there is not a reason^{E} to get "35". Since the fully rational agent would get the answer "35", the necessary condition is met.

As well, since it is restricting it to rational possibility, it seems likely that the scope of IP has been restricted. The concern that an agent's physical limitations would nullify the existence of a reason^{E} might not be relevant, since IP might not be requiring that it be physically possible

[22] This is the interpretation of "possibility" as understood by Korsgaard, "Skepticism, 11; Smith, "Internal", 112; Shafer-Landau, 172.
[23] However, Finlay rejects the rational possibility interpretation. See "Obscurity", p. 8.

for an agent to act on a putative reason to φ. It is only that it must be possible for an agent to act on it, insofar as he is fully rational.

Since there is this alternative reading, my argument that the interrelationship principle should be interpreted as being about reasons^{H} is not decisive. However, IPAR still has more initial plausibility as an interpretation of Williams than IPAN. But given this alternative, that claim needs additional support. In what follows, I will aim to provide that support.

The correct formulation of Williams's argument against external reasons. To show that the IPAR interpretation is the most plausible interpretation of IP, I will now construct an argument against external reasons, one that is intended to reflect Williams's thinking. Williams claims that the fundamental motivation for internalism is the interrelationship between explanatory and normative reasons. Therefore, the general version of the interrelationship principle, IP, will serve as one of the foundations of the argument. Using the general version of the interrelationship principle will allow us to plug in the various specific interpretations of it and consider the implications of them for the soundness of the argument against external reasons.

In addition to IP, one additional premise is needed for the argument. Without it, the argument would not go through. IP itself does not say anything about the elements in an agent's S. So it will not establish the internalist claim that a lack of a motivational element in the agent's S nullifies the existence of a normative reason. The content of the additional premise can be extracted from Williams's description of externalism as the claim that "it can be true that A has a reason [here we interpret this phrase more generally as a "normative reason"] to φ without there being any shadow or trace of that [motivation] presently in his S"[24]. In opposition, the internalist claim then is that there *cannot* be a normative reason if there is not *presently* a related motivational element in A's S. And the reason for the internalist claim, on Williams's thinking,

[24] Williams, "Internal Reasons", 39.

is that "when the reason is an explanation of his action, then of course it will be, in some form, in his S, because certainly – and nobody denies this – what he actually does has to be explained by his S."[25] What Williams is denying is that it is possible for some reason *r* to be part of an explanation of an agent's φ-ing if the agent does not presently have a motivational element in his S related to *r*. Put positively, he is claiming that it is possible for a reason *r* to be part of an explanation of an agent's φ-ing only if there is presently a motivational element in the agent's S. So, we have the following. (It is labeled PS-because it states that there must *presently* be an element in the agent's *S*.)

PS If it is possible for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.[26]

In PS I have added "correctly" to account for Williams's claim that we are allowed to correct an agent's beliefs in determining the normative reasons of an agent. By saying that the element in S must be correctly related to *r*, it rules out instances in which an agent would be motivated to φ for reason *r*, but only because the agent has a false belief. Without this qualification, the argument would not support the internalist requirement of there being a *sound* deliberative route to the agent's motivation to φ.

Williams's scenario of the agent who is motivated to mix what is in the glass with tonic and drink it is a good example of the rationale for the "correctly related" requirement. Let us suppose for this example that desires generate reasons to act. In this case, since the agent desires

[25] Ibid.
[26] Some philosophers, e.g. Michael Smith, will reject premise 3 as a component of internalism. Smith takes internalism as a claim about the hypothetical desires of an agent, not his actual desires. In Chapter 3 I will explain why that interpretation is incorrect.

to drink what is in the glass, it seems that the desire generates a normative reason *R* to φ (to mix what is in the glass and drink it). However, although there is an element in S related to *R* (the agent's desire to φ), the element is not *correctly* related to *R* because the agent is motivated to act on *R* only because he falsely believes that what is in the glass is gin. In other words, there is an element in the agent's S related to *R*, but because that element only exists due to a false belief of the agent, *R* is not a normative reason for A. Therefore, the agent's desire to drink what is in the glass, *R* (or perhaps the *putative* reason *R*), is not actually a normative reason for A to φ—even though we have supposed for this example that desires generate reasons^{E} to act.

PS is a very controversial claim. In fact, most of the objectors to internalism and/or its relativity claim will reject it. It might appear that the argument is now begging the question against externalism. But that's not the case. PS does not claim that there is a normative reason only if there is an element in the agent's S related to the reason. PS is not equivalent to the internalist claim. For one, it is merely the claim that it is possible for *r* to be part of an explanation of A's φ-ing only if there is an element in A's S related to the reason. Second, it does not state any necessary conditions for the existence of a normative reason. It plays a role in denying a normative reason only once it is combined with IP (as we will soon see). And, of course, whether PS is plausible will depend in large part on the notion of "possibility" being used within it. That issue will be addressed later.

Although eventually it will be appropriate to be concerned with the truth of PS, it is important to keep in mind that at present we are not concerned with whether the argument against external reasons is sound. Therefore the truth of PS is largely irrelevant. We are only concerned with formulating an argument based on IP which can support the internalist position. It may be that IP and any other premises necessary to support that conclusion are false. In fact,

what I am aiming to show is that it is only on the reasons^{H} interpretation that the premises of the argument are all true (without presupposing a quasi-instrumental theory of practical reason). On the reasons^{E} interpretations, one or more of the premises *will* be false, unless a quasi-instrumental theory is presupposed.

IP and PS are the basis for Williams's argument against the possibility of external reasons. From them we can derive a valid argument against the possibility of external reasons. Because they serve as the foundation for the argument against external reasons, I will label them the *foundational premises.* The argument against external reasons is a *reductio.* Since it is, we will assume the existence of an external reason in order to show that that it results in a contradiction. From Williams's statement in IROB about the nature of externalism, we can derive the following statement of the existence of an external normative reason.

ER There is a normative reason *r* for A to φ and there is not presently an element in A's S correctly related to *r.*

We now have all of the necessary components of the argument against external reasons. It is as follows.

IP If there is a **normative reason** *r* for A to φ, then it is **possible** for A to φ for reason *r*.

If it is **possible** for A to φ for reason *r*, then it is **possible** for *r* to be part of an explanation of A's φ-ing.

PS If it is **possible** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.

TP1. There is a **normative reason** *r* for A to φ and there is not an element in A's **S** correctly related to *r.*

2.	There is a **normative reason** for A to φ.	1; simplification
3.	It is **possible** for A to φ for reason *r*.	IP, 2; modus ponens
4.	It is **possible** for *r* to be part of an explanation of A's φ-ing.	IP, 3; modus ponens
5.	There is an element in A's S related to *r*.	PS, 4; modus ponens
6.	There is not an element in A's S related to *r.*	1; simplification
C.	If there is not an element in A's S correctly related to *r*, then *r* is not a **normative reason** for A to φ. (~TP1)	5,6; contradiction

The bold portions of the argument indicate the terms which are open to alternative interpretations. TP1, which is just the statement that there is an external reason, is assumed for *reductio*. We can now see that *if* IP and PS are true, it does follow that there are no external normative reasons. Of course we have to ask, are 1-3 actually true? Whether they are true, or at least whether certain philosophers will believe them to be true, will depend on how the bold terms are interpreted.

As we have already seen, the bold portions are open to various interpretations. "**Normative reason"** can be interpreted as either reasons^{E} or reasons^{H}. "**Possible**" can be interpreted in four ways. It can mean *logical, physical, rational, or "within an agent's capacity".* We have not considered the logical or physical possibilities yet, primarily because they are not held by any of the interpretations of internalism. However, Shafer-Landau considers them before

he ultimately dismisses them as plausible interpretations of Williams, and so we will examine them in Chapter 3.[27] Although there are various interpretations of the argument above, we will not consider most of them in this chapter. In the rest of this chapter I am only concerned with showing that the reasons^{H} interpretation of Williams's internalism is *pro tanto* plausible insofar as it renders the argument against external reasons sound. In Chapter 3 we will consider the other interpretations of the argument and their implications for it.

3. The argument against external reasons is sound on the reasons^{H} interpretation

As I stated above, in the rest of this chapter I am only concerned with showing that the reasons^{H} interpretation is *pro tanto* plausible insofar as it renders the argument against external reasons sound without presupposing a theory of practical reason. On this interpretation of Williams, "normative reason" is interpreted as "has a reason" and "possible" is interpreted as "within an agent's capacity". So the two premises for the argument against external reasons are as follows.

IPAR^{W}	If **A has a reason** *r* to φ, then it is **within A's capacity** to φ for reason *r*. If it is **within A's capacity** to φ for *r*, then it is **within A's capacity** for *r* to be part of an explanation of A's φ-ing.
PS^{W}	If it is **within A's capacity** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.

And the specific conception of an external reason on my interpretation is the following:

[27] Shafer-Landau, *Moral*, 172-3.

ER^H A **has a reason** *r* to φ and there is not presently an element in A's S correctly related to *r*.

The conclusion of the argument against external reasons becomes:

C If there is not presently an element in A's S correctly related to *r*, then A does not **have a reason** to φ.

We have already seen that the argument against external reasons is valid, and specifying the meaning of its ambiguous terms does not alter that fact. But is it sound on my interpretation? That depends on whether $IPAR^W$ and PS are both true. That $IPAR^W$ is true should be noncontroversial. And that is because the definition of a $reason^H$ is that it is one that an agent is capable of acting upon. To have a reason to φ it must be within the agent's capacity to act on the reason *r*. And if it is within the agent's capacity to act on reason *r*, then it must be within A's capacity for *r* to be part of an explanation of A's φ-ing. If we deny it, we would have to be committed to the following claim: An agent can φ for reason *r*, and yet *r* can have no part in the explanation of A's φ-ing. That seems logically impossible. If an agent is motivated by a particular reason to actually φ, that reason is necessarily part of the explanation of the agent's φ-ing.

The truth of PS^W might at first seem to be doubtful; however, ultimately it is not. There are two likely causes of doubt. The first is that it may appear to presuppose an instrumental or quasi-instrumental conception of practical reason. It might appear to be claiming that an agent

can come to be motivated to perform an action only if he already has a desire related to the action. However, S is not being construed narrowly so as to include only an agent's desires. S is being construed as widely as possible so as to include *any* component of an agent's psychology that can motivate him to act. This includes desires, beliefs, dispositions of evaluation, and so on. If an agent can be motivated by Kant's categorical imperative, for example, then that motivational capacity is an element in the agent's S. This wide construal of S precludes any objections by those who think that PS^W denies non-instrumental or desire-based forms of practical reasoning. Whatever types of practical reasoning a particular agent recognizes and/or engages in can be included as an aspect of the agent's S. However, as already noted, to be included in the agent's S it must be that the agent in question can actually be motivated by that type of reasoning.

The second concern might be with the inclusion of "correctly". The "correctly" qualification requires that an agent's capacity to be motivated to φ for *r* must not rest upon a false belief. It might be objected that even if the motivation to φ for *r* rests upon a false belief, that does not entail that the agent is therefore incapable of φ-ing for reason *r*. Even though it rests upon a belief, the result is that the agent is capable of φ-ing for reason *r*.

We should reject this idea (as does Williams).[28] Suppose that after a long day working in the sun, a man decides to swim in a lake to cool off. Unbeknownst to him, alligators inhabit the lake. Let us represent the action of swimming in a lake *un*inhabited by alligators with ψ, and the action of swimming in a lake inhabited by alligators as φ. The reason in this case, to cool off, will be *C*. So, in this case, the man falsely believed that he was going to ψ for reason C. Instead,

[28] On p. 107 of IER, Williams implicitly claims that internalism is concerned with the actions which an agent can *intentionally* perform, as he writes that "nothing can explain an agent's (intentional) actions except something that motivates him to act". The parenthetical is his. Although Williams thinks we should reject this idea, at present we are concerned with whether the premise is true, not whether it is consistent with Williams's thought. It may be consistent with his thought, but it could be false. So, we must determine whether it is true.

he φ-ed. Since he was motivated by reason *C*, and (unintentionally) φ-ed as a result, should we say that he was capable of φ-ing for reason *C*? I do not think we should. Had he known that the lake was inhabited by alligators, he would not have been motivated by *C* to swim in the alligator-infested lake.

However, some might continue to protest. It could be argued that in this case it *was* within the agent's capacity to φ for *C.* The agent φ-ed (swam in the alligator infested waters), and it was for *C* (to cool off). It just so happens that he did so unintentionally—and the first premise does not require that it be within the agent's capacity to *intentionally* φ for *C*. If someone wants to hold this position, then we should merely amend the first premise to state that "it must be within the agent's capacity to *intentionally* φ for reason *r*". The only reason to reject this modification to the interrelationship principle is if you think that an agent can have a reason to do something even if it requires the agent to be deceived in order for him to do so. This seems to be an implausible position since we are not concerned with reasons the reasons which it would be *epistemically* rational to act upon—that is, what considerations it would make sense for an agent to act upon given his current beliefs (whether true or false).

I do not think that we should make the modification to the first premise. However, for anyone who thinks that an agent's ability to accidentally φ for reason *r* satisfies the first premise as literally stated, then they should understand it to include the intentional qualification. Given Williams's insistence that internalism is concerned with what an agent would be motivated to do if the agent did not have false beliefs, if including the "intentional" qualification is necessary for the argument against external reason to be sound, I think Williams would be more than amenable to that modification.[29]

[29] This seems to be further supported by his claim on page 107 of IER that "nothing can explain an agent's (intentional) actions except something that motivates him to act." He recognizes that agents can perform actions

We have now examined each of the foundational premises of the argument against external reasons on the reasons^{H} interpretation. Since they are true (or would be with a slight modification to PS^{W}), the argument against external reasons is sound on the reasons^{H} interpretation. That makes the reasons^{H} interpretation charitable with respect to Williams's argument, and so gives us some reason to think that it is the correct interpretation of Williams. The strength of my argument will be significantly increased once we see in Chapter 3 that the other interpretations do not result in a sound argument, at least without presuming a particular theory of practical reason.

Conclusion

Two major objectives were accomplished in this chapter. First, I laid out in detail the reasons^{H} interpretation of Williams's internalism. According to the reasons^{H} interpretation, internalism is the requirement that in order for an agent to have a reason to φ it must be within the agent's capacity to be motivated to φ for that reason. We have seen that when we correct for an agent's beliefs, we are doing so in order to determine what the agent's motivations actually are (undistorted by false beliefs), and not what they would be were the agent to be fully practically rational. Also, we saw that internalism's construal of the agent's motivational capacity is very liberal. It is not restricted merely to the agent's desires. Any aspect of an agent's psychology that can motivate him to act is a constituent of his motivational capacity.

The second objective was to provide three considerations that support the accuracy of the reasons^{H} interpretation. I showed first that the reasons^{H} interpretation is the only interpretation

---

unrelated to their motivations, but only unintentionally so. He is concerned with the actions which an agent can intentionally perform.

which is consistent with both (R) and (N). Second, it provides the most plausible interpretation of Williams's interrelationship principle. And third, the reasons^{H} interpretation is plausible because it renders Williams's argument against external reasons sound, and without presupposing a particular theory of practical reason. Given these three considerations, there is good reason to think that it is the correct interpretation of Williams. However, given that there are alternative interpretations of Williams's internalism, further defense of my interpretation is needed. In the next chapter I will be concerned with showing why the reasons^{H} interpretation is to be preferred to the alternatives.

Chapter 3: A Defense of the Accuracy of the Reasons^{H} Interpretation

Before I explain the objectives of this chapter, let me recap what has been accomplished so far in the first two chapters. On the whole, I have been building a case for the claim that the reasons^{H} interpretation of Williams's internalism is the correct interpretation. On the reasons^{H} interpretation, internalism is the claim that an agent *has* a reason *r* to φ only if it is within the agent's capacity to be motivated by *r* to φ. The case for the reasons^{H} interpretation being the correct interpretation rests upon four considerations which, when taken together, constitute a strong argument in favor of the reasons^{H} interpretation. The first three considerations were established in Chapters 1 and 2, and are as follows. First, only the reasons^{H} interpretation is consistent with Williams's claims that, (R), all reasons for action are relative to an agent's subjective motivational set, and (N), no theory of practical reason is presupposed by internalism. The second consideration is that the reasons^{H} interpretation provides the most plausible interpretation of the interrelationship principle. And the third consideration is that the reasons^{H} interpretation renders Williams's argument against external reasons sound, and does so without presupposing a theory of practical reason.

The first objective of this chapter, which will be the subject of the first section of the chapter, is to establish the truth of the fourth consideration. The fourth consideration is that the reasons^{H} interpretation is the *most charitable* interpretation with respect to Williams's argument against external reasons. Since in Chapter 2 I already showed that the reasons^{H} interpretation renders the argument sound (without presupposing a particular theory of practical reason), all that must be accomplished in the first section of this chapter is to show that the other

interpretations are problematic in some way with respect to the argument against external reasons. If they are, they will then be less charitable interpretations of Williams's internalism.

The other interpretations that we will consider will be problematic with respect to Williams's argument against external reasons in one of three ways. The interpretations will 1) render the argument unsound, or 2) render the argument sound, but only by presupposing a particular theory of practical reason, or 3) be inconsistent with the argument. The latter would occur if the interpretation implicitly or explicitly rejects one (or more) of the premises of the argument against external reasons.[1] An interpretation which results in the philosopher's position and his argument for it being consistent with each other is, *ceteris paribus*, more charitable than one which does not.

If I am correct that each of the interpretations suffers from one of these problems, then the reasons^{H} interpretation will be the most charitable interpretation with respect to Williams's argument against external reasons. I will then have established the four considerations which support the reasons^{H} interpretation. So, there will be very good reason to think that the reasons^{H} interpretation is the correct interpretation of Williams's internalism. However, there might be some worries that the reasons^{H} interpretation conflicts with some portions of Williams's writing on internalism.

Therefore, the second objective of this chapter, which will be in the second section, is to respond to some possible objections to the accuracy of the reasons^{H} interpretation. Although in section one I will have completed a strong argument for the accuracy of the reasons^{H} interpretation, some philosophers will likely be concerned that the reasons^{H} interpretation is inconsistent with significant passages of Williams's account of internalism. In order to respond

[1] This is not same as "1" because, though the interpretation *claims* that one (or more) of the premises is false, it may be that the premise(s) are true (since the interpretation could make a false claim), and so the argument is actually sound.

to these possible objections, I will consider the passages of Williams which I think are most likely to appear inconsistent with the reasons^{H} interpretation. What I will argue is that most of the passages do not actually conflict with the reasons^{H} interpretation. To show that, I will explain how the passages allow for a different reading or understanding than the one which conflicts with the reasons^{H} interpretation, a reading that is consistent with it.

If both objectives are met, we will have conclusive reason to think that the reasons^{H} interpretation is the correct interpretation of Williams's internalism. In order to reach that conclusion, let us begin with the claim that the reasons^{H} interpretation is the most charitable interpretation with respect to Williams's argument against external reasons.

I. Charity and interpretation: Williams's argument against external reasons

This section is concerned with the fourth consideration in favor of my claim that the reasons^{H} interpretation of Williams's internalism is the correct interpretation: the reasons^{H} interpretation is the most charitable interpretation with respect to Williams's argument against external reasons. To show that it is the most charitable, we will consider five different interpretations of Williams's internalism. To begin with, there are two interpretations which we did not cover in Chapter 1, but which merit consideration.[2] Shafer-Landau briefly considers these interpretations—before rejecting them—in his book *Moral Realism*. Shafer-Landau characterizes internalism as the claim that there is a reason^{E} for an agent to φ only if it is possible for the agent to be motivated to φ. In trying to determine the truth of internalism, Shafer-Landau considers what notion of possibility is in use in the internalist claim. Two of those notions are *logical*

[2] I did not mention these interpretations in Chapter 1 because no philosophers (that I am aware of) accept these interpretations or consider them plausible interpretations of Williams.

possibility and *physical possibility*. (There is also a third notion—which is the notion which Shafer-Landau thinks Williams is using—which we will consider later.) On the logical possibility interpretation, internalism is the claim that reasons^{E} to act are constrained by logical possibility. A reason^{E} to φ can exist only if it is logically possible for an agent to be motivated to φ. On the *physical possibility interpretation*, internalism is the claim that reasons^{E} to act are constrained by physical possibility. A reason^{E} to φ can exist only if it is physically possible for an agent to be motivated to φ.

The last three interpretations to be considered are ones which we already saw in Chapter 1. They are the most plausible alternatives to the reasons^{H} interpretation. In Chapter 1 claimed that there are two predominant interpretations of Williams's internalism. The first is that *reasons^{E} to act are constrained by the subjective motivational set of the agent*. The second is that *reasons^{E} must be capable of motivating fully rational agents.* For the sake of easy reference, I will call the latter interpretation the *fully rational interpretation*. With respect to the first interpretation, however, I noted that there are two different understandings of the basis for the internalist claim. The most common understanding (held by Russ Shafer-Landau, Brad Hooker, Jay Wallace, and Rachel Cohon, amongst others) is that internalism presupposes a quasi-instrumental account of practical reason. I will call this the *quasi-instrumental interpretation*. The other understanding, held by Stephen Finlay, is that Williams is relying upon a novel conception of a "reason for action". I will call this the *novel conception interpretation.* This disagreement over the basis for internalism is essentially a disagreement over the nature of Williams's argument against external reasons. Because in this chapter we are concerned with the implications of each interpretation for Williams's argument against external reasons, we will need to evaluate these two different understandings as separate interpretations. Therefore, for

our present purposes, there are *three* predominant interpretations to be evaluated. Combined with the logical and physical possibility interpretations, in total there are five interpretations to be considered.

One further and very important point about these interpretations needs to be made. Despite their differences, all of the interpretations share one thing in common: their understanding of the type of reason for action which internalism is making a claim about. All of them take internalism to be making a claim about reasons^{E} to act (i.e., reasons that there are). That of course distinguishes them from the reasons^{H} interpretation of internalism. As should be clear now, they are also different from the reasons^{H} interpretation in that they do not accept the "within an agent's capacity" interpretation of "possible".

As I have said, part of the objective of this section is to show that each of the interpretations result in Williams's argument being unsound, being sound but only by presupposing a particular theory of practical reason, or being inconsistent with the argument. To demonstrate this, for the most part we will consider the implications of each interpretation for the formulation of William's argument against external reasons which was developed in Chapter 2 (which is also restated below). Since the argument was formulated at a level general enough so as to allow for it to be consistent with most interpretations of internalism, we can evaluate most of the five interpretations by inputting their particular content into the argument. However, the interpretation by Finlay is not conducive to being plugged into the formulation I developed. But, thankfully, Finlay has provided enough information to get a good idea of how he understands Williams's argument against external reasons. So, we can evaluate his interpretation with respect to his own understanding of the argument.

Before we consider the implications of each interpretation for Williams's argument, let us now quickly revisit the general formulation of Williams's argument against external reasons that was developed in Chapter 2.[3] The argument is as follows.

IP If there is a **normative reason** *r* for A to φ, then it is **possible** for A to φ for reason *r*.

If it is **possible** for A to φ for reason *r*, then it is **possible** for *r* to be part of an explanation of A's φ-ing.

PS If it is **possible** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.

TP1. There is a **normative reason** *r* for A to φ and there is not an element in A's **S** correctly related to *r*.

2.	There is a **normative reason** for A to φ.	1; simplification
3.	It is **possible** for A to φ for reason *r*.	IP, 2; modus ponens
4.	It is **possible** for *r* to be part of an explanation of A's φ-ing.	IP, 3; modus ponens
5.	There is an element in A's S related to *r*.	PS, 4; modus ponens
6.	There is not an element in A's S related to *r*.	1; simplification
C.	If there is not an element in A's S correctly related to *r*, then *r* is not a **normative reason** for A to φ. (~TP1)	5,6; contradiction

In Chapter 2 I demonstrated that this argument is valid. Whether it is sound depends upon the precise meaning of the ambiguous terms "normative reason" and "possibility". In Chapter 2 I showed that if we interpret the former as "reason^{H} (a reason that an agent *has*)" and the latter as

[3] My defense that this is an accurate formulation of his argument can be found in Chapter 2, p. 20-33.

"within an agent's capacity", then the argument is sound. What I will show below is that the result of specifying the two terms in accordance with the other interpretations (Finlay's interpretation excluded), the interpretations will render the argument unsound, render it sound only by presupposing a particular theory of practical reason, or be inconsistent with the interpretation. We will examine the five interpretations in the following order: 1) The logical possibility interpretation, 2) the physical possibility interpretation, 3) the quasi-instrumental interpretation, 4) the novel conception interpretation, and lastly, 5) the fully rational interpretation. Let us now examine the first interpretation.

1. The logical possibility interpretation

As I mentioned above, the logical possibility interpretation comes to us via Shafer-Landau (though he rejects it as the correct interpretation of Williams's internalism). Shafer-Landau characterizes internalism as the claim that there is a reason^{E} for an agent to φ only if it is possible for the agent to be motivated to φ. He then considers that claim on the interpretation of "possible" as logical possibility. The resulting interpretation of internalism is that it is the claim that there is a reason^{E} for an agent to φ only if it is *logically possible* for the agent to be motivated to φ.

So, what are the implications of this interpretation for Williams's argument against external reasons? To see that, we can begin by noticing that this interpretation provides the specific content for the two ambiguous terms ("normative reason" and "possible") in the formulation of Williams's argument given above. "Normative reason" is understood as reason^{E}, and "possible" is understood as logically possible. We can now input that content into the entire

argument. However, for brevity's sake, I will only list the foundational premises (IP and PS) and the conclusion of the argument. Since the soundness of the argument depends on the truth of the foundational premises, all that needs to be shown is that IP or PS is false, and the argument will be shown to be unsound. That is what will be shown.

Since this interpretation interprets "normative reason" as reason^{E}, we can specify IP a bit more by substituting "reason^{E}" in for "normative reason". In Chapter 2 we saw that this is an *agent-neutral* conception of normativity[4], and so we labeled the interrelationship principle on that interpretation IP-AN (Interrelationship Principle-Agent Neutral). Since the notion of possibility which is in use is logical possibility, I will add the superscript "LP" to both of the foundational premises. Lastly, the conclusion C will become "C-AN" to indicate that the agent-neutral notion of normativity is in use.

Given the above changes, the foundational premises and conclusion are the following. (The "↓" indicates the intermediate premises have been left out.)

IP-AN^{LP}	If there is a **reason^{E}** *r* for A to φ, then it is **logically possible** for A to φ for reason *r*.
	If it is **logically possible** for A to φ for *r*, then it is **logically possible** for *r* to be part of an explanation of A's φ-ing.
PS^{LP}	If it is **logically possible** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.
	↓
C-AN	If there is not an element in A's S correctly related to *r*, then *r* is not a **reason^{E}** for A to φ.

[4] See Chapter 2, II.2.

So what are we to make of this argument? Since logical possibility is the widest notion of possibility, this might seem to be a reasonable interpretation of Williams, given that Williams did not specify which notion of possibility he was using.

However, we must consider what the implications of that interpretation are for the argument. Does it make the argument sound? No. And that is because PS^{LP} is false. PS^{LP} claims that only if there is a motivational element in the agent's subjective motivational set which is correctly related to *r* is it *logically* possible for *r* to serve as part of an explanation of the agent's φ-ing. Let us put this in negative form, as I think it makes the nature of the claim a little clearer. PS^{LP} claims that if an agent does not presently have a motivational element correctly related to reason *r,* then it is logically impossible for *r* to be part of an explanation of the agent's φ-ing.

To see why this is false, it must first be noted that the antecedent of PS^{LP} does not specify that *r* must be part of an explanation of A's φ-ing *at the present moment*. That is, it does not say that, in order for it to be logically possible for an agent to φ at t_1, the agent must have a motivational element related to *r* at t_1. (That would make PS^{LP} more plausible.[5]) Instead, it claims that for there to be a $reason^{E}$ *r* to φ (perhaps now, but also perhaps later), the agent must presently have a related element in his S.

Since *when* the agent is to φ for reason *r* is left open, as long as we can come up with a logically possible future scenario in which *r* is part of an explanation of A's φ-ing (which might necessarily involve the agent acquiring a motivational element related to *r*) despite A's not presently having a motivation related to *r*, then we will have shown PS^{LP} to be false. And that is rather easy to do. It could be that an agent who presently has no motivation to φ for reason *r* has

[5] However, there is good reason to think that Williams did not intend to restrict it to the present moment. That is because Williams allows that there can be a normative reason for an agent to φ even if the agent must deliberate in order to come to be motivated to φ. And deliberation takes time.

his brain tinkered with by mad scientists, causing him to acquire the motivation to φ for reason *r*. Though far-fetched, the scenario is certainly logically possible, and so PS^{LP} is false.[6] It *can* be logically possible for *r* to be part of an explanation of an agent's φ-ing even if presently there is not an element in the agent's S related to *r*.[7] Since PS^{LP} is false, the argument against external reasons is unsound on this interpretation.

Since the logical possibility interpretation of internalism renders the argument unsound, it is less charitable than the $reasons^{H}$ interpretation (with respect to Williams's argument against external reasons). In addition, there is also another reason to think that this is not Williams's view. It seems that Williams would almost certainly have recognized that an agent's current motivational profile would have little, if any, impact on the *logical* possibility of a reason being part of an explanation of the agent's action. Therefore, charity requires dismissing this as a plausible interpretation of Williams.

2. The physical possibility interpretation

The next interpretation interprets "possibility" as *physical* possibility. In this case "physically possible" is being used in its most general sense: what is possible given the laws of physics. In using this notion, we are not concerned with what some particular agent is physically capable of performing, in the way that we can be concerned with whether an agent is physically capable of jumping across a river. *That* notion of possibility would be more akin to the $reasons^{H}$ interpretation of "within an agent's capacity", but specific to the agent's physical capacity. Instead, we are concerned with what is possible, given the laws of physics.

[6] This scenario is given by Sobel in "Explanation, Internalism, and Reasons for Action", *Social Philosophy & Policy* 18, no. 2 (2001): 222.
[7] Shafer-Landau makes this point in *Moral Realism*, p. 171.

Like with the logical possibility interpretation, we will modify the premises, this time in accordance with the physical possibility interpretation. Since the notion of possibility is *physical possibility*, the superscript "PP" has been added to IP-AN and PS. The notion of "normative reason" is again "reason^{E}". The premises and conclusion on the physically possible interpretation are the following.

IP-AN^{PP}	If there is a **reason^{E}** *r* for A to φ, then it is **physically possible** for A to φ for reason *r*. If it is **physically possible** for A to φ for *r*, then it is **physically possible** for *r* to be part of an explanation of A's φ-ing.
PS^{PP}	If it is **physically possible** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.
	↓
C-AN	If there is not an element in A's S correctly related to *r*, then *r* is not a **reason^{E}** for A to φ.

So what about this argument? Is it sound? Again, no. In this case both the truth of IP-AN^{PP} and PS^{PP} can be called into question. We will take a look at the problems for both, though PS^{PP} will be our central concern. With respect to IP-AN^{PP}, we might think that there can be a reason^{E} to perform an action even if the laws of physics prevent it. If I am outside of a baseball stadium and I do not have a ticket, nor can I see in to watch the game, it seems that there could be a reason^{E} for me to hover 10 feet off of the ground so that I can see over the fence and watch the game; and this is so despite the fact that, given the laws of physics, it is impossible for me to

hover. The desire I have to see the game, or the pleasure it would bring me to watch, and so on, seem to generate a reason^{E} to hover above the fence. So, I think IP-AN^{PP} is doubtful. However, whether it is true might depend on our account of practical reason, and so I cannot say definitively that it is false.

But PS^{PP} is false. PS^{PP} claims that it is physically possible for a reason *r* to be part of an explanation of an agent's φ-ing only if there is presently a motivational element in the agent's S correctly related to *r*. Let us put this in negative form, again for the sake of clarity. PS^{PP} claims that if an agent does not presently have a motivational element correctly related to reason *r* then it is physically impossible for *r* to be part of an explanation of the agent's φ-ing. Why should we think that whether it is physically possible—i.e. possible given the laws of physics—for an agent to be motivated to φ for reason *r* depends upon the agent *presently* having a motivational element related to *r*?

There are two good reasons to think that it is not. For the first reason, keep in mind what I pointed out earlier, that the antecedent of PS^{PP} does not specify that *r* must be part of an explanation of A's φ-ing *at the present moment*. Future actions are included in its scope as well. Like with the logical possibility version of PS, to show that PS^{PP} is false, we just need to come up with a future scenario which is physically possible—i.e. within the laws of physics—in which *r* is part of an explanation of A's φ-ing, despite the agent's not presently having a motivation related to *r*. As Shafer-Landau points out, even if there is not an element in an agent's S which is correctly related to *r*, the laws of physics do not prevent the agent from *acquiring* the motivation to φ for reason *r*.[8] To illustrate his point, we can again use the mad scientist scenario. It is possible within the laws of physics for mad scientists to meddle with an agent's brain in order to cause them to have a desire to perform an action for a certain reason. Or, we can also use a

[8] Shafer-Landau, *Moral Realism*, 172.

normal scenario, one not from science fiction. Someone who has not acquired the taste for beer may have no motivation to drink a beer (φ) for the pleasure (*r)* of it. But the agent might acquire the taste for beer. Perhaps he is pressured by his friends to drink beer with them. For the reason that it will help him fit in with his friends (which is a different reason from pleasure) he may be motivated to drink beer. And that can lead to his acquiring the taste for beer. If so, *then* he could drink beer (φ) for the sake of pleasure (r). Undoubtedly this is physically possible—within the laws of physics. Therefore, PS^{PP} is false, and so the argument against external reasons is unsound. That is enough to show that the physical possibility interpretation of internalism is less charitable than the reasons^{H} interpretation (with respect to Williams's argument against external reasons).

So, both the logical and physical possibility interpretations have been shown to be less charitable than the reasons^{H} interpretation. Let us now move on to the predominant interpretations which were introduced in Chapter 1.

3. The quasi-instrumental interpretation

This interpretation is the most common interpretation of Williams's internalism. It is the first of two interpretations which take internalism to be the claim that reasons^{E} to act are constrained by the subjective motivational set of an agent. The second of the two interpretations is Stephen Finlay's *novel conception interpretation.* As I mentioned in Chapter 1, what distinguishes these two interpretations from each other is their understanding of the *basis* for internalism's claim that reasons^{E} to act are constrained by an agent's motivational set. As can be discerned from their names, the former thinks that internalism relies upon a quasi-instrumental

theory of practical reason, whereas the latter thinks it relies upon a novel, evidentialist, conception of "reason for action" (which amounts to a novel theory of practical reason). Given that both of these interpretations claim that internalism relies upon a particular theory of practical reason, it will not be surprising that they render Williams's argument against external reasons sound, *but* only on the presupposition of a particular theory of practical reason. However, we still need to do the work of showing that their interpretations do have that implication for Williams's argument. In this sub-section we will focus only on the quasi-instrumental interpretation.

The two most important aspects of the quasi-instrumental interpretation are its understandings of the notions "normative reason" and "possibility" in the formulation of Williams's argument against external reasons which I provided in Chapter 2. Like all of the interpretations considered in this section, it understands "normative reason" to be reasons^{E}. But with respect to "possibility", it understands internalism to be using the notion of *rational* possibility. We can now modify the foundational premises to reflect this interpretation, adding the superscript "RP" to the relevant premises. On this interpretation, the foundational premises and conclusion become:

IP-AN^{RP} If there is a **reason^{E}** *r* for A to φ, then it is **rationally possible** for A to φ for reason *r*.

If it is **rationally possible** for A to φ for *r*, then it is **rationally possible** for *r* to be part of an explanation of A's φ-ing.

PS^{RP} If it is **rationally possible** for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S correctly related to *r*.

↓

C-AN If there is not an element in A's S correctly related to *r*, then *r* is not a **reason**^{E} for A to φ.

To evaluate the soundness of the argument, we must first determine what is meant by "rationally possible". With that understanding in hand, we will then consider, one at a time, the truth of IP-AN^{RP} and PS^{RP}. As we will see, it is the truth of PS^{RP} which will be called into question.

So, what does it mean to say that something is rationally possible? It will help to first note that, in this context, we are concerned with actions, and, in particular, whether it is rationally possible for an agent to perform an action, i.e., to φ. What does it mean to say that it is rationally possible for an agent to φ? To say that φ-ing is rationally possible for an agent is to say that the agent would φ *were he fully rational*. That is, were he fully rational, he would choose to perform the action. To be rationally possible, it is not relevant what action the current agent, i.e., the actual agent, would perform. Instead, what is relevant is what action the agent *would* perform, *if* he were fully rational.

That many philosophers have come to this understanding of the notion of possible is due to Williams's claim that, in determining what there is normative reason for an agent to do, we can correct for the rationality of the agent. So, since we can correct an agent's reasons for their rationality, it seems as though we are concerned *not* with what reasons the agent-as-he-is is motivated by, but instead with what reasons he would be motivated by were he fully rational.[9]

But what does it mean for an agent to be fully rational (at least in this context)? According to the quasi-instrumental interpretation, it means that the agent is fully *practically* rational. An agent who is fully practically rational will have not only (at least) all true beliefs

[9] In the second section of this chapter I will explain why I think this is a flawed interpretation of "possible".

relevant to the situation which he is in[10], but also, *if* some motivations are rationally required, those motivations which are required by reason. Instrumentalists and Aristotelians, amongst others, disagree on whether motivations can be rationally required. Instrumentalists say no, while Aristotelians say yes. According to the former, what is rational for an agent to do depends on whatever motivations the agent just happens to have. No motivations are in themselves in accordance with, or contrary to, reason. According to the latter, however, some motivations are intrinsically required by reason. For example, if an agent lacks courage (which is a motivational state of the agent), then he is not fully practically rational. Whether motivations, and perhaps especially more narrowly desires, can be required by reason is certainly an ongoing debate. But the phrase "fully practically rational" does not, in itself, presume in favor of any theory of practical reason. Whether being fully practically rational requires having particular motivations depends on whatever turns out to be the correct theory of practical reason.

Given that we have a more specific understanding of the notion of "rationally possible", it will help to modify IP-AN^{RP} in accordance with it. Unfortunately, given that "rationally possible" is understood as "what an agent would do were he fully practically rational", merely substituting the latter in for the former would result in an incoherent sentence. So we are going to have to finagle the premise a little bit. What we get is the following. (I have added "-M" to the superscript to indicate that it is the modified version.)

[10] Michael Smith points out that being fully practically rational should probably not require having *all* true beliefs. That is because sometimes what there is reason^{E} for an agent to do is affected by his current lack of true beliefs. For example, if I do not know what time the bus is leaving, there is (presumably) a reason^{E} for me to obtain a brochure with the bus times listed. But, if we required that a fully rational agent have *all* true beliefs, then he would already know what time the bus is arriving, and so he would not obtain a brochure. But since what there is reason for the less-than-fully practically rational agent to do is determined by what the fully rational agent would do, there would not be a reason^{E} for the less-than-fully rational agent to obtain a brochure. That seems to be the wrong conclusion.

IP-AN^{RP-M} If there is a **reason^{E}** *r* for A to φ, then **the fully practically rational version of A would** φ for reason *r*.

If **the fully practically rational version of A would** φ for reason *r*, then *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing.

So, is IP-AN^{RP-M} true? It seems so. Some philosophers have noted that the first sentence appears to be a platitude or tautology.[11] It is merely the claim that a genuine reason^{E} for action is one which a fully practically rational agent would act upon. And the second sentence seems even less dubitable. It merely claims that a reason which was the basis for a fully practically rational agent's action would serve as part of an explanation of the agent's action. How could it not?

Although questions can be raised about the truth of IP-AN^{RP}, I will not bog us down here with a detailed defense of IP-AN^{RP-M}.[12] That is unnecessary for our purposes. Remember that the reason we are examining the quasi-instrumental interpretation and its implications for Williams's argument against external reasons is to show that the argument on that interpretation is

[11] Respectively, Finlay, "Obscurity" 8; Shafer-Landau, *Moral Realism*, 173. It should be noted that what I am claiming Shafer-Landau says is tautologous is the notion of internalism which he understands Korsgaard to accept. However, that notion is equivalent to IP-AN^{RP}. What Shafer-Landau will object to is the conception of rationality (which he thinks, but I do not) Williams is implicity relying upon (quasi-instrumental rationality).

[12] I actually have my own doubts about the truth of IP-AN^{RP} as specifically stated here, but I will not pursue that concern. That is because PS^{RP} will make the soundness of the argument on this interpretation rely upon a particular theory of practical reason, which is enough to show that this interpretation of Williams's internalism is less charitable than the reasons^{H} interpretation. My concern with IP-AN^{RP} is that it appears to "silence" all other reasons^{E} which there might be for acting in a particular situation. Presumably, a fully rational agent would perform the action which there is *most* reason^{E} to perform. But, since IP-AN^{RP} claims that what there is reason^{E} to do is determined by what the fully practically rational agent would do, according to IP-AN^{RP} there would be no reason^{E} *at all* to perform any actions which the fully rational agent would not perform. But that seems incorrect. There can be plenty of reasons^{E} to perform various actions in a particular circumstance even if they are not the one which the fully rational agent would perform. Even if a fully rational agent would study for an exam, and not watch television the night before the exam, surely the pleasure which would result from watching television is *a* reason^{E} to do so. However, faced with this objection to IP-AN^{RP}, those who accept it could easily modify it to the claim that what there is *most* reason^{E} to do is determined by what a fully rational agent would do. That, in fact, is more in line with Williams's thinking. See "Postscript", 91.

problematic. We can show that it is by *presuming* IP-AN^{RP-M} to be true. That is because, when we conjoin it with PS^{RP}, it will become clear that the truth of PS^{RP} depends upon a quasi-instrumental theory of practical reason. So, let us move on to PS^{RP}.

Before we see why the truth of PS^{RP} depends upon a quasi-instrumental theory of practical reason, let us get a firmer grip on our understanding of PS^{RP}. In the original formulation of Williams's argument, PS was merely the claim that if it is possible for *r* to be part of an explanation of A's φ-ing, then there is presently an element in A's S related to *r*.[13] As I stated in Chapter 2, PS is essentially denying that it is possible for an agent to perform an action for a reason *r* if they do not have any motivation properly related to *r*. On the reasons^{H} interpretation, where "possible" is interpreted as "within an agent's capacity", that is not a controversial claim. But that is not the case on the quasi-instrumental interpretation of "possible" as rational possibility.

Like we did with IP-AN^{RP}, we will modify PS^{RP} to reflect the specific meaning of "rationally possible" as "what an agent would do were he fully rational". And, also like with IP-AN^{RP}, we will have to finagle the wording a little. (Again, I add "-M" to the superscript.)

PS^{RP-M} If *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing, then there is presently an element in A's S correctly related to *r*.

Given the rationally possible interpretation of "possible", the claim of PS^{RP-M} is that, if *r* would be part of an explanation of the fully practically rational version of A's φ-ing, then A (the actual and *less-than-fully* rational agent) must *presently* have an element in his S correctly related

[13] See Chapter 2, II.4.

to *r*. To put it in negative form, $PS^{RP\text{-}M}$ claims that if the actual (less-than-fully rational) agent A does *not* presently have a motivational element correctly related to *r*, then *r* would not be part of an explanation of the fully rational version of A's φ-ing. To see why $PS^{RP\text{-}M}$ depends upon a quasi-instrumental theory of practical reason, let us put both foundational premises together.

$IP\text{-}AN^{RP\text{-}M}$	If there is a **reason^{E}** *r* for A to φ, then **the fully practically rational version of A would** φ for reason *r*. If **the fully practically rational version of A would** φ for reason *r*, then *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing.
$PS^{RP\text{-}M}$	If *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing, then there is presently an element in A's S correctly related to *r*.

Again, we are going to presume that $IP\text{-}AN^{RP\text{-}M}$ is true. To see that *if* it is, $PS^{RP\text{-}M}$ depends upon quasi-instrumentalism, consider a scenario in which an agent has no motivation to perform an action. To demonstrate this, it will probably help if we pick an action which non-instrumentalists might think there is a reason^{E} to perform, no matter whether the agent is motivated to perform it. Suppose that Dan has no motivation to go to work and he knows that if he does not show up he will get fired. And, living paycheck to paycheck, he knows that being fired will result in his not being able to pay rent and so he will likely be kicked out of his apartment.

In this case, the putative reason for Dan to go to work is that he will likely lose his apartment if he does not go. Let us call this reason *L*. Even when Dan thinks through the scenario

and realizes that he will likely lose his apartment if he does not go to work, he still has no motivation (*at all*) to go to work. He does not have an element in his S related to *L*. According to $PS^{RP\text{-}M}$, what follows from this? Given that the lack of a motivational element means that the consequent of $PS^{RP\text{-}M}$ is false, by *modus tollens* it follows that *L* would *not* be part of an explanation of the fully practically rational version of Dan's φ-ing. But, also by *modus tollens,* it then follows that the antecedent of the second conditional of $IP\text{-}AN^{RP\text{-}M}$ is false; the fully practically rational version of Dan would not φ for reason *L*.

Although there is one further logical implication which we will consider, let us stop here briefly to see the importance of our current conclusion. What we have just seen is that according to $PS^{RP\text{-}M}$, what actions the fully practically rational version of an agent would perform is constrained by the motivations of the actual (less-than-fully practically rational) agent. But which theories of practical reason would accept that claim? Only theories which think that what it is rational to do depends *entirely* on an agent's motivations. That is, only instrumentalist or quasi-instrumentalist (given that we have already seen that "motivations" on Williams's account are construed more broadly than just desires) theories of practical reason. Already this is enough to see that the truth of $PS^{RP\text{-}M}$ depends upon a quasi-instrumental theory of practical reason.

But let us look at the last implication of Dan's complete lack of motivation to go to work. Given that the fully practically rational version of Dan would not φ for reason *L*, it follows that *L* is not a $reason^{E}$ for Dan to φ (again by *modus tollens*). In other words, the fact that Dan will likely lose his apartment if he does not go to work is not a $reason^{E}$ for him to go to work. So, what there is $reason^{E}$ for Dan to do depends entirely upon his present motivations. We can now

see why Shafer-Landau, Parfit, MacIntyre, and so on reject internalism.[14] They think that there can be reasons^{E} to act unrelated to an agent's existing motivations.[15]

As I mentioned above, the purpose of considering each of these five interpretations of Williams's internalism is to show that each interpretation results in one of the three following problems with respect to Williams's argument against external reasons: 1) rendering the argument unsound, 2) rendering the argument sound, but only by presupposing a particular theory of practical reason, or 3) being inconsistent with the premises of the argument (i.e. implicitly or explicitly claiming that one of the premises is false). As we have just seen, the quasi-instrumental interpretation results in problem #2. It renders the argument sound, but only by presupposing a particular theory of practical reason (quasi-instrumentalism). Therefore, the quasi-instrumental interpretation of internalism is less charitable than the reasons^{H} interpretation.

4. The novel conception interpretation

The novel conception interpretation is held by Stephen Finlay. As I mentioned previously, the novel conception interpretation agrees with the quasi-instrumental interpretation in understanding internalism to be the claim that reasons^{E} to act are constrained by the subjective motivational set of an agent. But it disagrees with the quasi-instrumental interpretation on the *basis* for the internalist claim. As I mentioned previously, that is essentially a disagreement over the nature of Williams's argument against external reasons. Therefore, since what we are concerned with in Section I is the impact of the different interpretations of Williams's

[14] Shafer-Landau, *Moral Realism*, 185; Parfit, "Reasons and Motivation", 130; MacIntyre, *Dependent Rational Animals*, (Chicago: Open Court, 2005), 86-7.
[15] We can also see why those who accept a quasi-instrumental conception are likely to accept internalism. They would think that the reasons^{E} to act of an agent are determined entirely by their existing motivations. They would likely accept PS^{RP}, and so also agree with internalism.

internalism on his argument against external reasons, we need to evaluate the novel conception interpretation separately from the quasi-instrumental interpretation. But before we evaluate the implications of the novel conception interpretation for Williams's argument, let us first get a refresher on the nature of the interpretation.

As we saw in Chapter 1[16], Finlay claims that Williams's concept of a normative reason for action is: "*R* is a reason for A to φ' means that *R is an explanation of why A would be motivated to φ if he deliberated soundly*."[17] Finlay calls this an evidentialist model of practical reason.[18] It is not an instrumental or even quasi-instrumental conception of practical reason. It is not claiming that an agent can only reason about how to satisfy his desires. Rather, it is claiming that a reason for action must be such that the agent for whom it is a reason must be able to recognize or "see" the reason as being one on which he would act were he correctly informed. And the evidential model claims that if an agent can see himself acting on a putative reason^{E}, there must be a motivational element in his S which he sees as being related, in some way, to the putative reason^{E}. For example, suppose that the local college is putting on a performance of "The Nutcracker". Let us call that fact *N*. If *N* is going to be a reason^{E} for me to attend the performance, according to the evidential conception of a reason^{E} for action, I will have to see myself as being motivated by *N* to attend the performance. But, if I absolutely despise the ballet and have *no* motivation to see the ballet, then I would not see myself as being motivated by *N* to φ. Therefore, *N* is not a reason^{E} for me to attend the ballet.

So how does this understanding of the basis for internalism affect Williams's argument against external reasons? To determine that, we will have to operate differently than we did with the previous interpretations. The nature of Finlay's interpretation resists the substitution of its

[16] See Chapter 1, III.2.B.
[17] Finlay, "Obscurity," 14.
[18] Ibid., 16.

content into the formulation of the argument which I developed (at least in any clear way). So, instead of plugging the relevant concepts into the argument like we did with the previous interpretations, in this case we will rely on Finlay's own (inchoate) formulation of the argument. According to Finlay, the first premise in Williams's argument against external reasons is the evidentialist conception of "reason for action".

> 1. '*R* is a reason for *A* to *φ*' means that *R is an explanation of why A would be motivated to φ if he deliberated soundly*.

An interesting aspect of this premise is that it not only provides (what Finlay takes to be) the internalist conception of a reason for action, but it also implicitly contains the interrelationship principle. It maintains that there is a necessary relationship between normative reasons, *R*, and explanatory reasons (since *R is* an explanation). So, Finlay's argument captures Williams's claim that the interrelationship between normative and explanatory reasons is a fundamental motivation for internalism. Unfortunately, however, Finlay is not clear on what the other premises are.

But the provision of the first premise is enough to determine that the novel conception interpretation of Williams's internalism is problematic. As is obvious, the first premise is the statement of a particular conception of "reason for action". Such a conception constitutes a particular theory of practical reason, i.e. a theory of what generates a reason^{E} for action. Thus, we can see that, even if the rest of the premises of the argument (whatever they may be) are true, the soundness of the argument depends upon a particular theory of practical reason.[19] So, the

[19] That it depends upon a particular theory of practical reason might seem to be especially problematic in this case given that the theory of practical reason is novel, and as Finlay points out, even *radical* (see p. 13). Why would

novel conception interpretation of internalism suffers from problem #2. Even assuming that all of the other premises are true (since we do not know what they are), the novel conception interpretation renders Williams's argument against external reasons sound, but only by presupposing a particular theory of practical reason. Therefore, the novel conception interpretation of internalism is less charitable than the reasons^{H} interpretation with respect to Williams's argument against external reasons.

Although it goes outside my general objective in Section I (showing that other, non-reasons^{H} interpretations are problematic with respect to Williams's argument against external reasons), let me provide one further reason to think that the novel conception interpretation is not the correct interpretation of Williams's internalism. That additional reason is that Williams denies that his theory relies on an evidentialist model (though Williams did not state it in those terms). Finlay claims that on Williams's model of deliberation, "the elements in an agent's motivational set ('*S*') appear in the 'foreground', as the content of the agent's reasoning."[20] That is, to be a reason^{E} for action, the agent must see the reason as being one that he would act upon, *and* he has to see the reason as being one he would act upon *because* it is related to his S.

Williams explicitly rejects this model. In "Postscript" he writes:

> Its [a reason for action's] making normative sense to him implies that it made normative sense in terms of his S. This does not mean that when an agent has a thought of the form "that is a reason for me to *φ*," he really has, or should really have, the thought "that is a reason for me to *φ in virtue of my S*." The disposition that forms part of his *S* just is the

---

Williams think that others would be convinced to accept internalism on the basis of a *novel* conception of practical reason? However, Finlay takes Williams's writing to not only be providing an argument for internalism, but also an argument for his conception of what it is to be a reason^{E} for action. So, if Williams is successful with the latter, he could also be successful with the former.

[20] Ibid., 16.

> disposition to have thoughts of the form "that is a reason for me to φ," and to act on them.[21]

For Williams, our motivational set affects what putative reasons we are motivated by. But in being motivated by reasons (due to our S), we do not have to think of the reason being constituted by its relation to our S. In stark contrast to Finlay's claim, the elements of the agent's S are (or at the very least, can be) in the background of the agent's practical deliberation, not the foreground. Therefore Finlay's novel conception interpretation is inconsistent with Williams's account of internalism, and so should be rejected.

Before we move on to the last interpretation to be considered, let me address one possible objection that might be raised by holders of the quasi-instrumental or novel conception interpretations of internalism. As we saw, both of those objections resulted in Williams's argument against external reasons being sound only on the presupposition of a particular theory of practical reason. I have claimed that this is problematic (though not necessarily fatally problematic) for an interpretation of Williams's internalism. The holders of those interpretations might respond by pointing out that they would have expected the soundness of Williams's argument to depend upon the presupposition of the truth of a particular theory of practical reason. That it does presuppose a theory of practical reason (one which they reject) is why they have rejected internalism. So, it could be claimed that my demonstration that their interpretation has that result does not show that their interpretation is incorrect—since they would have expected it to have that result.

However, that they would have expected such a result does not mean that the result does not count as a problem for their interpretation. Given that Williams has denied that internalism

[21] Williams, "Postscript", 93.

rests upon a particular theory of practical reason, an interpretation which renders his argument for that position sound without presupposing a particular theory of practical reason is, *ceteris paribus*, more charitable than one which does not. Hence, the reasons^{H} interpretation is more charitable than either the quasi-instrumental or novel conception interpretations, *with respect to Williams's argument against external reasons*. There could be other considerations which override this consideration and therefore provide us with most reason to accept a non-reasons^{H} interpretation (though I do not think there are). But, with respect to Williams's argument against external reasons, the reasons^{H} interpretation is more charitable than both the quasi-instrumental and novel conception interpretations.

5. The fully rational interpretation

The last interpretation that we have left to evaluate interprets internalism as the claim that reasons^{E} for action must motivate fully rational agents. Or, in other words, if a putative reason^{E} to act is one which would not motivate a fully rational agent to perform the action, then it is not a reason^{E} to act. Both Christine Korsgaard and Michael Smith accept this view. What is most distinctive about this view is that it does not take (R), the claim that all reasons are relative to an agent's motivational set, to be essential to the internalist position. That is, they think that one can accept internalism, and yet not also accept the claim that all reasons^{E} for action are relative to the agent's motivational set. Whether the relativism claim is true depends on the correct account of practical reason. And, as they understand it, internalism is not wedded to any particular theory of practical reason. For that reason, despite the fact that they reject the relativism claim, they accept internalism. However, because the fully rational interpretation does not take the relativism claim

to be essential to internalism, it suffers from problem #3 with respect to Williams's argument against external reasons—the interpretation is inconsistent with the argument. In particular the fully rational interpretation is inconsistent with Williams's argument in that it rejects the foundational premise PS.

Like the first three interpretations in this section, the fully rational interpretation can be plugged into the formulation of Williams's argument originally developed in Chapter 2. Like the quasi-instrumental interpretation, the fully rational interpretation interprets "normative reason" as *reason*^{E} and "possible" as *rationally possible*. Also, they both agree on the understanding of "rationally possible" as "what an agent would do were he fully practically rational". And, so, at least initially, we appear to get the same version of the argument as on the quasi-instrumental interpretation (though we will see that the versions of the argument turn out to be substantially different).

IP-AN$^{RP\text{-}M}$ If there is a **reason**$^{\mathbf{E}}$ *r* for A to φ, then **the fully practically rational version of A would** φ for reason *r*.

If **the fully practically rational version of A would** φ for reason *r*, then *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing.

PS$^{RP\text{-}M}$ If *r* would be part of an explanation of **the fully practically rational version of A's** φ-ing, then there is presently an element in A's S correctly related to *r*.

↓

C-AN If there is not an element in A's S correctly related to *r*, then *r* is not a **reason**$^{\mathbf{E}}$ for A to φ.

However, although those who accept the fully rational interpretation—namely, Korsgaard and Smith—will agree with the terminology of the interpretation which has been plugged in, they will reject premise $PS^{RP\text{-}M}$. In particular, they will reject the claim that whether a particular reason *r* can be part of an explanation of a fully practically rational agent's action depends upon whether there is presently a motivational element in the (less-than-fully practically rational) agent's S related to *r*. By claiming that they would reject this premise I do not mean that they think it is false (although they do), but instead I mean that they would not accept it as a premise which is essential to the argument against external reasons (and so for internalism). That is because they reject the idea that internalism is concerned with the motivations of the agent-as-he-is. Instead, they take it to be concerned with the motivations of the fully practically rational agent.

As evidence that they would reject $PS^{RP\text{-}M}$, here is Smith's rejection of Williams's relativistic claim (which we previously saw in Chapter 1).

> Now in fact it is initially quite difficult to see why Williams says any of this [referring to a couple of quotes by Williams stating the relativistic thesis] at all. For, as we have seen, what the internalism requirement suggests is that claims about an agent's reasons are claims about her *hypothetical* desires, not claims about her *actual* desires. The truth of the sentence 'A has a reason to φ' thus does not imply, not even 'very roughly', that A *has* some motive which will be served by his φ-ing; indeed A's *motives* are beside the

> point…What the internalism requirement implies is rather that A has a reason to φ in certain circumstances C just in case he would desire that he φs in those circumstances if he were fully rational.[22]

The most important aspect of this quote is that in the first couple of sentences we see that Smith separates the relativism claim from internalism. The relativism claim, on his understanding of internalism, is not essential to it. One can accept internalism while at the same time rejecting the relativity of reasons claim.

And why does he think the relativism claim is not an essential part of internalism? Because internalism, as he understands it, is not concerned with the motivations of the current (less-than-fully practically rational) agent. Instead, it is concerned with the motivations of the fully practically rational agent. And, what motivations the fully practically rational version of the agent will have is not necessarily tied to the less-than-fully practically rational agent's motivations. Therefore what reasons the fully rational version of an agent would be motivated by (and so would be part of an explanation of the agent's actions) may not have any relation to the less-than-fully rational agent's motivations. What reasons the fully rational version of an agent would be motivated by is necessarily tied to the less-than-fully practically rational agent's motivations *only if* we *presume* an instrumental or quasi-instrumental theory of practical reason. That is, there is such a necessary connection only if we presume that what actions are rational depends entirely on an agent's existing motivations. (Though expressed in a slightly different manner, Korsgaard agrees with Smith's position with respect to the relativistic thesis.[23]) So, according to Smith (and Korsgaard), the relativity thesis is not essential to internalism, and so

[22] Smith, "Internal Reasons", 117.
[23] Korsgaard also claims that, for internalism, the motivations of the less-than-fully rational agent are relevant *only if* we accept a particular theory of practical reason. See "Skepticism", 19-23.

PS^{RP-M} is not a part of the argument against external reasons, as it would make the relativity thesis essential to internalism.

A bit of an aside: One crucial thing to keep in mind here is that because it is *Williams's* internalism with which we are concerned, it is Williams's prerogative to determine the nature of the internalist position. If he wants the relativity thesis to be essential to internalism, then it is. If he also wants to claim that the relativity thesis does not depend on any particular theory of practical reason, then that claim is also part of the internalist position. Of course, although it is his prerogative to determine the nature of the internalist position, it does not follow that the internalist position is correct. Williams gets to determine what the position is, but not whether it is true.

For example, if I decide to make up a theory, let us call it "Nonsense", it is my prerogative to determine which theses are essential to it. Suppose that I claim Nonsense is composed of two essential theses: A) Only dogs are lovable animals, and B) Garfield is a lovable animal. Obviously the position of Nonsense is self-contradictory. But that does not mean that A and B are not essential to Nonsense. Even if someone points out to me that A and B are contradictory, he or she does not get to say that therefore B is not actually a part of the theory of Nonsense. What he can say is that the position of Nonsense is self-contradictory.

Why do I say this? I say it because although Korsgaard and Smith are aware that Williams claims the relativity thesis is a part of internalism, they think that it should be thrown out.[24] However, it is not Korsgaard or Smith's prerogative to throw it out. If Williams says that essential to internalism are both 1) the claim that all reasons to act are relative to an agent's S, and 2) the claim that internalism is not committed to any particular theory of practical reason, then they are essential to internalism. They may be inconsistent, but that does not allow

[24] Korsgaard, "Skepticism", 19-23. Smith, "Internal", 117.

Korsgaard or Smith to remove one of them from the internalist position. Of course, if Korsgaard and Smith want to come up with *their own* position, and call it internalism, then they can do so. (But it would be nice if they pointed out that their internalism was an alternative to William's internalism, and not intended to be the same position as Williams.) The aside is now complete.

Given that Korsgaard and Smith's fully rational interpretations of Williams's internalism deny that the relativism claim is essential to internalism, it is clear that the fully rational interpretation is problematic with respect to Williams's argument against external reasons. It results in problem #3: the interpretation is inconsistent with the argument against external reasons. Since Smith and Korsgaard do not think that the relativism thesis is a necessary part of internalism, but would only follow from the presumption of an instrumental theory of practical reason, they would reject the inclusion of $PS^{RP\text{-}M}$ in Williams's argument against external reasons. And that is because $PS^{RP\text{-}M}$ would make the relativism thesis an essential aspect of internalism. However, as I showed in Chapter 2[25], PS (not specifically $PS^{RP\text{-}M}$, as the latter only results from Korsgaard and Smith's particular interpretation of internalism) *is* a part of Williams's argument. Therefore, the fully rational interpretation of internalism is less charitable than the $reasons^{H}$ interpretation with respect to Williams's argument against external reasons.

We have now seen that all five interpretations in this section are problematic with respect to Williams's argument against external reasons. All of them are problematic in one of three ways: 1) rendering the argument unsound, 2) rendering the argument sound, but only by presupposing a particular theory of practical reason, or 3) being inconsistent with the argument. Given that the $reasons^{H}$ interpretation avoids all three of those problems, with respect to Williams's argument against external reasons, it is the most charitable interpretation.

[25] See Chapter 2, II.4.

The positive case for the reasons^{H} interpretation is now complete. In particular, it has now been shown that there are four considerations which, taken together, strongly support the claim that the reasons^{H} interpretation is the correct interpretation of Williams's internalism. First, only the reasons^{H} interpretation is consistent with Williams's claims that, (R), all reasons for action are relative to an agent's subjective motivational set, and (N), no theory of practical reason is presupposed by internalism. Second, the reasons^{H} interpretation provides the most plausible interpretation of the interrelationship principle. Third, the reasons^{H} interpretation is charitable insofar as it renders Williams's argument against external reasons sound, and does so without presupposing a theory of practical reason. And fourth, the reasons^{H} interpretation is the *most charitable* interpretation with respect to Williams's argument against external reasons.

However, before we conclude that the reasons^{H} interpretation is the correct interpretation of Williams, we need to consider some possible objections to the accuracy of that interpretation.

II. Objections to the accuracy of the reasons^{H} interpretation

Despite the fact that the four considerations above strongly support the conclusion that the reasons^{H} interpretation is the correct interpretation of Williams's internalism, I suspect that there will be some resistance to that conclusion. In particular, I think that some people will claim that some passages by Williams—ones which are pivotal to his account of internalism—conflict with the reasons^{H} interpretation. So, they might think that a more reasonable conclusion is that, although the reasons^{H} interpretation is the most plausible with respect to those four considerations, overall, due to its conflict with important passages of Williams's accounts of

internalism, it should be rejected. One possibility, after all, is that Williams's account of internalism is self-contradictory. So, perhaps there is no consistent interpretation of Williams.

In order to respond to those objections and show that the reasons^{H} interpretation is not inconsistent with Williams's account of internalism, in this section we will take a look at several passages of Williams which might seem to conflict with it (and possibly even support one of the alternative interpretations rejected above). What I will argue is that the passages do not actually conflict with the reasons^{H} interpretation. To demonstrate that, I will explain how the passages allow for a more plausible alternative reading than the one which conflicts with the reasons^{H} interpretation, a reading that is consistent with it. If I am successful, that will give us sufficient reason to think that the reasons^{H} interpretation does not conflict with the correct interpretation of Williams's internalism. And, given the four considerations in favor of the reasons^{H} interpretation, we will have conclusive reason to accept the reasons^{H} interpretation as the correct interpretation of Williams's internalism. We now turn to the passages which appear to conflict with the reasons^{H} interpretation.

1. The claim that an agent can be unaware of a normative reason

The first passage by Williams which might seem to conflict with the reasons^{H} interpretation comes from his initial account of internalism in "Internal and External Reasons" (IER). What this passage will seem to contradict is the reasons^{H} interpretation's understanding of both "normative reason" as "reason^{H}" and "possible" as "within an agent's capacity". And that is because the passage will seem to indicate that there can be a normative reason for an agent even though it is not within the agent's capacity to act on it. But by definition, for an agent to

have a reason to φ (a reason^{H}), it must be within the agent's capacity to act upon the reason. Therefore, *if* Williams claims that there can be a normative reason for an agent to φ which it is not within the agent's capacity to act on, then by "normative reason" he must *not* mean reason^{H}, but some other type of reason; and by "possible" he must *not* mean "within an agent's capacity", but some other type of possibility. I will argue that such a reading of Williams only results from a less-than-careful reading of the text, and that upon closer examination the passage actually supports the reasons^{H} interpretation.

So why might it seem that Williams claims there can be a normative reason for an agent to φ even if it is not within the agent's capacity to φ? That is because, according to Williams, since internalism is concerned with an agent's rationality[26], an agent:

> may not know some true internal reason statement about himself…and [this] comes from two different sources. One is that *A* may be ignorant of some fact such that if he did know it he would, in virtue of some element in his *S*, be disposed to φ; we can say that he has a reason to φ, though he does not know it….The second source is that *A* may be ignorant of some element in *S*.[27]

So why does this seem to conflict with the interpretation of "possible" as "within an agent's capacity"? If there can be a normative reason for an agent to φ, and yet the agent is ignorant of the fact which would motivate him to φ, then it seems as though it is not within the agent's capacity to perform that act. For example, suppose that a coffee shop which I often frequent is

[26] As I mentioned in Chapter 1, "rationality" is ambiguous between theoretical rationality and practical rationality. The former is concerned with only the beliefs of the agent, whereas the latter is possibly, depending on the correct theory of practical reason, concerned both with the beliefs of the agent as well as the desires. See pages 29-30.
[27] Williams, "Internal", 103.

giving away free coffee in honor of their 25th year of business, but that I am unaware of that fact. If I were aware of that fact, I would be motivated to go to the coffee shop. It seems that, according to the quote above, Williams would say that there could be a normative reason for me to go to the coffee shop, since I would be motivated to go were I made aware of the fact that they are giving away coffee. However, because I am unaware of it, it seems that it is not within my capacity to act on the reason. Were they to send out an email alerting me of the promotion, or a friend were to call me up to tell me about it, then I would go. But those two events are not within my capacity to bring about. So, it is not within my capacity to go to the coffee shop for the reason that they are giving away free coffee. Hence, it seems that Williams is using "normative reason" in a sense other than "reason^{H}".

However, despite appearances, that is not the case. A crucial sentence follows the above quote, one which I think many readers of Williams have missed. Immediately following the sentence which claims there can be a normative reason for an agent to φ even if the agent is unaware of the fact which would motivate him, Williams writes: "For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that *A* would have a reason to φ if he knew the fact"[28,29] (emphasis mine). What is crucial is that Williams is asserting that an agent's lack of knowledge *can* prevent there being a normative reason for the agent which there would otherwise be. Unfortunately, it is not clear what Williams means by "close and immediate". However, the idea seems to be that, for there to *actually* be a normative reason, it must be within the agent's capacity to learn the fact which would motivate him. If it is not within his capacity to learn the fact, then it is not *actually* a normative reason for him to act. This

[28] Ibid.

[29] It is not exactly clear what Williams means by the relevance of the fact being close and immediate. I think the best guess is that this is a less-thought-out idea of the "sound deliberative route" which he introduces in subsequent texts.

restriction on the agent's normative reasons, that he be aware of the fact which would motivate them, is not only consistent with the reasons^{H} interpretation, but it also seems to support it. Hence, this passage which at first glance might have appeared to conflict with the reasons^{H} interpretation turns out to provide evidence for it.

It is also worth noting that although this passage at first seemed to support the quasi-instrumental and fully rational interpretations of internalism, in fact it does not. In particular it seemed to support the "rationally possible" reading of "possible" with respect to the interrelationship principle. The normative reasons of an agent seemed to be determined by what action the agent would be motivated to perform, *were the agent fully practically rational*. However, since Williams claims that the lack of knowledge can sometimes preclude the existence of a normative reason, that indicates he is not relying on the "rationally possible" notion of "possible". Instead, although he does not state it explicitly, it appears that he is claiming that it must be within the agent's capacity to act on the reason--which obviously supports the "within an agent's capacity" reading of "possible" with respect to the interrelationship principle.

2. The claim that an agent can have a reason to φ, even if not currently motivated to φ

The next passage that we will look at appears to conflict with the idea that the motivations which an agent's normative reasons must be related to are the agent's *present* motivations. Instead, it seems that the motivations which internalism is concerned with are the ones which the agent would have if he were fully rational. Williams writes:

> It is important that even on the internalist view a statement of the form 'A has reason to φ' has *normative force*. Unless a claim to the effect that an agent has a reason to φ can go beyond what that agent is already motivated to do – that is, go beyond his already being motivated to φ – then certainly the term will have too narrow a definition. 'A has a reason to φ' means more than 'A is presently disposed to φ'. One reason why it must do so is that it plays an important part in discussions about what people should become disposed to do.[30]

If what an agent has reason to do can go beyond what he currently is motivated to do, does that not that show that internalism is *not* concerned with the agent's present motivations, but instead the motivations which his fully practically rational self would have? If internalism were concerned with the latter, that would support the "rationally possible" interpretation of "possible", and not the "within an agent's capacity" interpretation.

However, the claim that an agent's normative reasons can go beyond what he is currently motivated to do does not necessarily conflict with the "within an agent's capacity" interpretation. That is because, although an agent is not *at present* motivated to φ, that does not mean that it is not within the agent's capacity to be motivated to φ. In other words, it can be within an agent's capacity to φ even if the agent is not presently motivated to φ. In particular, it can be within an agent's capacity to φ, despite not presently being motivated to φ, by *acquiring* the motivation to φ.

How the agent is to acquire the motivation is seen in the formulation of internalism, which claims that an agent has a reason to φ only if there is a sound *deliberative* route from the agent's S to the agent's φ-ing. Although someone may not presently be motivated to perform an

[30] Williams, "Internal Reasons", 36.

action, he may be capable of deliberating about the action such that he comes to have the motivation to perform it. A high school student may not currently have the motivation to call a girl he likes, but, after some deliberation, he may acquire the motivation to call. Suppose at present he lacks the courage to call. Suppose his worry that she might say "no" prevents him from actually calling her. However, he could try to muster up the courage to call by engaging in some "deliberation". To get the courage needed to call, he might try to psych himself up, telling himself that he can do it, or he might remind himself that it is not the end of the world if she does say "no", and so on. Given this possibility, even if an agent is not presently motivated to φ, he can still have a reason to φ, since it is within his capacity to be motivated to φ. Since the reasons^{H} interpretation allows for an agent to have a reason to φ despite not presently being motivated to φ, there is not a conflict between the passage by Williams above and the reasons^{H} interpretation.

To further support my claim that the above passage does not conflict with the reasons^{H} interpretation, let me also address Williams's claim that internalism "plays an important part in discussions about what people *should* become disposed to do"[31] (emphasis mine). What is the sense of "should" here? We might naturally think that what someone should become disposed to do is what a fully rational person would become disposed to do. But that would be more in line with the rational possibility interpretation of the interrelationship principle (which conflicts with the reasons^{H} interpretation of the interrelationship principle).

However, we need to keep in mind that when we tell someone what he should become disposed to do, we can take two (if not more) different standpoints. One is the *ideal standpoint*. From it, we would be concerned with what the agent should be disposed to do, were he fully rational. The second is what we might call the *practical standpoint*. From it, we would be

[31] Ibid.

concerned with what the agent should *actually* be disposed to do given his limitations (one or more of which might be a deficiency in practical rationality and/or motivation). To put it another way, we could be concerned to tell an agent which action, of all of the actions which he is capable of performing, he should be motivated to perform.

The rational possibility interpretation takes the ideal standpoint. From it, statements about what an agent should become disposed to do are determined by what the agent would do were he fully rational. It is not concerned with whether the agent is actually capable of having the disposition—for the agent's lack of rationality may prevent him from acquiring it. For example, someone who suffers from an acute case of social anxiety disorder is likely, to that extent, deficient with respect to practical rationality. As a result, the agent lacks all sorts of dispositions to act which a fully practically rational agent would have. Whereas a fully practically rational agent would desire to go shopping, run errands, go to social events, and so on, the agent with social anxiety disorder does not. However, from the ideal standpoint we could say that the agent should come to have those motivations.

The "within an agent's capacity" interpretation takes the practical standpoint. With respect to what actions an agent should become disposed to do, it would be concerned with the dispositions an agent is actually capable of acquiring. For the agent who has social anxiety disorder, even though he is not presently disposed to perform any of the above actions, there may be some of them which it is within his capacity to become disposed to perform. Perhaps there is no way that he could bring himself to be motivated to go to a social event where he would be forced to engage in a lot of small talk. However, he might be able to develop the disposition to go grocery shopping. To gain that disposition, he may remind himself that people will largely leave him alone unless he asks for their assistance, that he can go through the "self-checkout"

lane (enabling him to avoid small talk with a cashier), and so on. If so, then from the practical standpoint we could tell him that he should gain the motivation to go shopping, but we could not truthfully tell him that he should become motivated to go to a social event.[32]

So, there are two different possible readings of Williams's claim that internalism plays a part in discussions of what an agent should become disposed to do. The text itself does not tell us which reading to take. At the least, however, there is a plausible reading of the text which is consistent with the reasons^{H} interpretation. So this text does not necessarily conflict with the reasons^{H} interpretation. And, given the whole of Williams's writings on internalism, it is more likely that Williams is taking the practical standpoint with respect to what an agent should become disposed to do. (This claim will receive even more support in the following sub-section, where we will see, in Williams's response to McDowell, Williams's claim that statements about what an agent should do on the internalist account are distinctively about the agent. That is, they are statements about what the agent should do, given his actual condition which includes any flaws the agent has. In other words, they are not statements of what the agent should do were he some idealized (fully practically rational) agent.) He is claiming that internalism plays a part in determining, of the dispositions which an agent is *capable* of acquiring, which ones he should acquire.

3. The claim that all reasons for action are internal

[32] In "Values, Reasons, and Persuasion" Williams aims to differentiate between scenarios where our advice to someone—telling them what they should do—helps them acquire the motivation for an action he already has reason to perform, and helping an agent acquire the motivation to perform an action which he does not already have reason to perform. In the former case, there is already a sound deliberative route from the agent's existing S to the action in question. So, in that case the one helping the agent does not change the agent's S, but merely helps them engage in the process of deliberation to arrive at the motivation. In the latter, there is not already a sound deliberative route from the agent's S to the action in question. Therefore, the agent does not at present have a reason to act. But, the "helper" may try to alter the content of the agent's S in some way so that he comes to have a reason to act. This could be done through persuasive speech, intimidation, and so on. With the example of the person with social anxiety disorder, statements of what they agent should become disposed to do are restricted to the former case.

The next, and final, possible conflict between the reasons^{H} interpretation of internalism and Williams's account of internalism is not found in a single passage. Rather, it is a group of statements scattered throughout Williams's writings on internalism. In multiple places Williams claims that *all* reasons for action are internal. In IER he writes that "the only real claims about reasons for action will be internal claims",[33] and "it is very plausible to suppose that *all* external reason statements are false"[34]; in IROB he states, "there are *only* internal reasons for action"[35]; and, similarly, in VRP he claims "the *only* reasons for action are internal reasons"[36] (the emphases in all of the above quotes are mine).

So why are these quotes a source of potential conflict with the reasons^{H} interpretation? They are troublesome because the reasons^{H} interpretation allows for two categories of reasons for action: reasons^{H}, which are all internal; but also reasons^{E}, which are *not necessarily* internal. Reasons^{E}, for all that the reasons^{H} interpretation says, *may* be external. That is, what there is reason for an agent to do (irrespective of whether the agent is capable of doing it) may not have any relation to the agent's existing motivations. But the quotes above seem to deny the possibility that there can be *any* reasons which are external, and so threaten the legitimacy of the reasons^{H} interpretation. To put it another way, the quotes above are problematic because, on the reasons^{H} interpretation of internalism, only reasons^{H} are within the scope of the internalist claim. But Williams's claim that *all* reasons for action are internal appears to imply that both reasons^{H} and reasons^{E} are within the scope of internalism.

[33] Williams, "Internal", 111.
[34] Ibid., 109.
[35] Williams, "Internal Reasons", 35.
[36] Williams, "Values", 109.

As the start of a response to the above statements, let me first say that I think Williams has been a bit careless in his wording. To say that all reasons for action are internal certainly suggests the literal reading that *all* reasons for action (both reasons^{H} and reasons^{E}) must be related to an agent's motivations. However, once we examine all of his writings, it will be clear that Williams did not intend to rule out the possibility of external reasons^{E}, but only external reasons^{H}. Several considerations show that Williams does at least allow for reasons^{E} that are not related to an agent's motivations.

The first consideration is found in Williams's response to T.M. Scanlon's concern about internalism. At issue is what we can say about a husband who is insensitive and/or cruel to his wife. Suppose that we provide him with several considerations which we think should motivate him to treat his wife nicer—e.g. pointing out the pain that his behavior causes her, mentioning that she deserves better given the nature of their relationship, and so on. According to internalism, if the cruel husband is not motivated by any of these considerations, then none of them provide a normative reason for the husband to treat his wife nicer. Although Williams says that we cannot (truthfully) say that there is a normative reason for the husband to treat his wife nicer, he says that it can still be "sensible" to say that the husband is "inconsiderate, or cruel, or selfish, or imprudent".[37] About this claim Scanlon writes:

> These criticisms…involve accusing [the agent] of a kind of deficiency, namely a failure to be moved by certain considerations that we regard as reasons. (What else is it to be inconsiderate, cruel, insensitive, and so on?) If it is a deficiency for the man to fail to see

[37] Williams, "Internal", 110.

> these considerations as reasons, it would seem that they must be reasons for him. (If they are not, how can it be a deficiency for him to fail to recognize them?)[38]

Scanlon's worry is that if, as internalism claims, an agent's lack of motivation to act on a putative normative reason entails that it is not a normative reason at all, then it seems internalism is denying that there can be normative reasons by which an agent fails to be motivated. To put it in a way relevant to our present discussion, Scanlon's worry is that internalism is denying that there can be a reason for action unrelated to an agent's motivations—i.e. an external reason.

That Williams did not intend to deny the existence of *any* external reasons can be seen in his response to Scanlon. Williams writes:

> I agree that the agent's faults can be understood in terms of a failure to see certain considerations as reasons…I also agree that if we think of this as a deficiency or fault of this man, then we must think that in some sense these reasons *apply* to him; certainly he cannot head off the criticism by saying that the reasons do not apply to him because he does not have that kind of *S*… But none of this implies that these considerations are already the defective agent's reasons; indeed, the problem is precisely that they are not.[39,40]

---

[38] T. M. Scanlon. *What We Owe to Each Other*. (Cambridge: Harvard University Press, 1999), p. 367. Quoted in Williams, "Postscript", p. 95.

[39] Williams, "Postscript", 95-6.

[40] Williams appears to contradict himself in this passage. On the one hand, he says that the reasons in question do apply to the defective agent; but, on the other, he says that these considerations are not the defective agent's reasons. How can they apply to the agent, but not be his reasons? Here is how I think we should understand him. When Williams says that they apply to the agent, what he means is that they are genuine considerations, i.e. reasons which count in favor of the agent acting in a particular way. They are reasons which, if the agent were fully rational, he would be moved by to act. But, since the agent is not fully rational—and so those considerations are not ones which are capable of motivating him as he is—they are not reasons that he has. If we interpreted internalism as being about reasonsE we would not have this resolution available to us, and so would leave Williams in a self-contradictory

Williams's admission that the agent is defective implies that Williams recognizes external reasons^{E}. When Williams says that the agent is deficient, with respect to what is he deficient? Given that the context of the writing is a discussion about reasons for action, the most plausible interpretation is that the agent is deficient with respect to practical rationality. And what could it mean to be deficient with respect to practical rationality other than that the agent is not motivated by the reasons which a fully practically rational agent would be motivated by? That is, there is a reason for action which, were he fully practically rational, he would be motivated by to act. Such a reason is an external reason^{E}.

The second consideration that favors interpreting Williams as allowing for some types of external reasons is his claim that on the internalist account statements about an agent's reasons for action are distinctively about the agent. He contrasts that type of reason statement with ones which are about what there is reason to do in particular circumstances.[41]

The distinction between reason-statements that are distinctively about an agent and reason-statements that are about what there is reason to do in particular circumstances is best illustrated in Williams's reply to John McDowell. In arguing for the possibility of external reasons, McDowell puts forth the idea that what normative reasons there are for an agent is determined by what reasons the *phronimos* (a fully practically rational agent) would be motivated to act on. In response, Williams claims that, on this view, statements that "A has a reason to φ" are not statements *distinctively* about A. That is because such statements "do not relate actions to persons, but types of action to types of circumstances, and they are most

---

position. That the reasons^{H} interpretation has the resources to resolve the apparent self-contradiction provides further support for it.

[41] Williams, "Values" 109; "Replies", 194.

revealingly expressed in the form 'in circumstances X, there is reason to φ'."[42] And, at the end of his response to McDowell, Williams goes on to say that internalism "is the only view that plausibly represents a statement about A's reasons as a distinctive kind of statement about, distinctively, A."[43]

A quick aside: It is worth noting that Williams begins the discussion about reasons for action with the locution "has a reason", but, in describing the externalist conception, he switches to "is a reason". Although he does not explicitly state that he is making a distinction between "is a reason" and "has a reason" in the same way that I have in this dissertation, this passage suggests that he is. But I am leery of resting my response on what may be an unintentional switch in locution.

So, why does Williams response to McDowell's argument support my claim that Williams does not intend to deny the possibility of external reasons^{E}? For one, after he notes that the externalist construal of "A has a reason to φ" is better understood as "In circumstances X, there is reason to φ", Williams does not object to the coherence or legitimacy of the latter type of claim (which is what many interpreters of Williams have thought he was doing in arguing for internalism). Rather, he responds to the latter claim by reiterating that, on the internalist account, statements about an agent's reasons are distinctively about the agent.

On the above externalist account, however, they are not distinctively about the agent. That is because what there is reason to do for an agent is determined by what the *phronimos* would do. Therefore every agent will have the same reasons to act, no matter how much their actual Ss vary. If the *phronimos* would be motivated to drink a beer in some circumstance, then there is a normative reason for every agent to drink a beer in that circumstance, no matter how

[42] Williams, "Replies", 190.
[43] Ibid., 194.

much their motivational profiles vary from the *phronimos*. So, an agent who can drink responsibly as well as a recovering alcoholic both have normative reason to drink a beer, even if having a beer will cause the recovering alcoholic to relapse.[44] Williams insists that the external conception comes to erroneous conclusions about the normative reasons for an agent, if those conclusions are distinctively about the agent (deficiencies and all).[45] In other words, according to Williams, external reason statements are problematic when they are put forth as though they are about what a particular agent, as he actually is, should do.

Also, if Williams really thought that *all* reasons for action were internal, it seems that he would have attacked the notion of identifying a reason for action in particular circumstances without knowing anything about the motivations of the agent in those circumstances. If he really though there are no external reasons^{E}, then he would not have thought that to be possible. To put it another way, it seems that he would have objected to the notion of determining what the *phronimos* would do in a particular situation. That is because, if there are no external reasons^{E}, there would be no *single* ideal *phronimos*. If all reasons are internal, then what there is normative reason for each agent to do would depend on his actual motivations (which Williams would allow to be "corrected" for false beliefs). So, what the fully practically rational version of each agent would do would vary from agent to agent. There would be a different *phronimos* for each agent. So, if Williams did not think that there are external reasons^{E}, it is reasonable to think that he would have argued against the coherence of the idea of a single ideal *phronimos*. But that is not what he did.

[44] Williams heads off a possible objection to this construal of the implications of the Aristotelian account. He writes that the Aristotelian cannot circumvent this problem "by putting all A's limitations into the account of the circumstances. If the circumstances are defined partly in terms of the agent's ethical imperfection, then the *phronimos* cannot be in *those* circumstances…". Williams, "Replies", 190.
[45] Ibid.

Instead, Williams thought it was important to establish that statements of reasons for action, as far as the internalist account is concerned, are distinctively about a particular agent—they say what an agent should do, given his actual constitution. Therefore, McDowell's objection misses the mark because it misunderstands the nature of the issue with which internalism is concerned. Internalism is *not* concerned with what action a supremely *ideal* agent would perform in a particular circumstance, but instead is concerned with what action an agent, given his actual make-up (i.e. physical, intellectual, and motivational capacities and so on) should perform. In other words, Williams's response to McDowell was not to deny the coherence of external *reason*^{E} statements, but rather to point out that internalism is concerned with reason^{H}-statements. (Reasons^{H} statements *are* distinctively about the agent, because they take into account the agent's actual motivational limitations.) So McDowell's objection is flawed because it takes internalism to be making a claim about reason^{E} statements, instead of reason^{H}-statements.

One last consideration supports my claim that Williams does allow for the possibility of external reasons^{E}. In "Values, Reasons, and the Theory of Persuasion", Williams addresses an objection by those who think that prudential and/or moral considerations are reasons by which any fully practically rational agent would be motivated. The objection is that since internalism allows us to correct for the rationality of an agent, we should be able to correct for the prudential and/or moral reasons by which the fully rational agent would be motivated. Williams responds:

> Now it may be claimed that prudential, or again moral, policies are similarly involved in what it is to be a fully rational agent, and some philosophers have claimed conclusions of this kind. It may, indeed, to some extent be true, particularly with regard to a modest amount of prudence; if an agent is totally devoid of a concern for the effects of actions on

> himself, we may indeed have problems in understanding what could count for him as a sound deliberative route at all. But to the extent that these things are true, then we are being told something about the necessary contents of the *S* of any rational agent.[46]

From this passage we can make an important inference: Williams allows for the possibility that there are some considerations by which any fully rational agent would be motivated; and that entails that there can be some reasons^{E} which are unrelated to a particular agent's motivational set.

To say that there are some motivational elements which are necessarily part of any fully rational agent's S must mean there are some considerations—at present not necessarily construed as *reasons*—by which any fully rational agent would be motivated. But, how could such *considerations* not be *reasons*^{E}? Surely they are. If they are not *reasons* to act, then it is not intelligible how it could be incumbent upon a *rational* agent to be motivated by them. Therefore, if there are some motivational elements which are necessarily a part of any fully rational agent's S, then that means there are some reasons^{E} by which any fully rational agent would be motivated.

But, if there are some reasons^{E} by which any fully rational agent would be motivated, that is inconsistent with the claim that *all* reasons for action (including reasons^{E}) are internal. To see why, we will begin by considering the implications of the latter claim. It says that all reasons^{E} (as well as reasons^{H}) are internal, i.e. they are all related to an agent's motivations in some way. Therefore, in determining what reasons^{E} an agent should be motivated by, we *must start with the motivations* of the agent, *and then* determine which considerations are reasons^{E} to act. Unless an agent has a motivation related to the consideration, it cannot be a reason^{E}. For example, we

[46] Williams, "Values", 111.

cannot say that the consideration "eating removes the pain of hunger" is a reason^{E} to eat unless we already know that a particular agent is motivated to avoid pain. If *all* reasons for action are internal, it is not possible to go the other way around. That is, we cannot first determine what constitutes a reason^{E} to act, and then determine that any fully rational agent would be motivated by it. No consideration can constitute a reason^{E} if there is not already a corresponding motivational element. But Williams's claim that there can be some motivational elements which are necessarily part of any fully rational agent's S *does* go the other way around. It implies that there are some considerations which constitute a reason^{E} to act independent of whether a particular agent is motivated by it or not.

Therefore, given that Williams allows for the possibility of determining some motivational elements which are necessarily part of any fully rational agent's S, he is implicitly denying that *all* reasons for action are internal. That is, he is implicitly allowing that there can be *some* reasons for action which are external. And, almost certainly he would claim that such reasons would be reasons^{E}, not reasons^{H}.

However, it could be argued that Williams has just contradicted himself. When he said that *all* reasons for action are internal, that is what he meant, and he just did not realize that his passage in VRP (quoted above) implied the external reasons^{E} are possible. If this was the only passage of Williams to rely upon in coming to an understanding of Williams, that would be a plausible position. But, given that the first two considerations we looked at above already indicate that Williams allows for the possibility of external reasons^{E}, there is more reason to think that he did *not* intend to claim that all reasons (including reasons^{E}) are internal. It is much more likely that he meant to claim that all *reasons^{H}* are internal.

We have now looked at the three passages which are most likely to appear to conflict with the reasons^{H} interpretation of internalism. What we have seen is that most often the apparent conflict is due to an ambiguity in the text, and that the ambiguity allows for a plausible alternative interpretation which is consistent with the reasons^{H} interpretation. There do not seem to be any good reasons for thinking that the reasons^{H} interpretation is inconsistent with Williams's various explanations of internalism.

Summary

So where are we now? Prior to this chapter I showed that there are three considerations which support the claim that the reasons^{H} interpretation of Williams's internalism is the correct interpretation. First, only the reasons^{H} interpretation is consistent with Williams's claims that, (R), all reasons for action are relative to an agent's subjective motivational set, and (N), no theory of practical reason is presupposed by internalism. The second consideration is that the reasons^{H} interpretation provides the most plausible interpretation of the interrelationship principle. And the third consideration is that the reasons^{H} interpretation renders Williams's argument against external reasons sound, and does so without presupposing a theory of practical reason.

In the first section of this chapter, I argued for the fourth consideration, that the reasons^{H} interpretation is the most charitable interpretation of internalism with respect to Williams's argument against external reasons. I did so by showing that the five other interpretations were problematic in some way (e.g. rendering the argument unsound, etc.). When the fourth consideration is combined with the first three considerations, that constitutes a very strong

argument for the conclusion that the reasons^{H} interpretation of Williams is the correct interpretation. In the second section of this chapter, I argued that the passages of Williams which might at first appear to conflict with the reasons^{H} interpretation do not. Hence, there is very strong reason to accept the reasons^{H} interpretation as the correct interpretation of Williams's internalism.

However, we are left with one further question. Is Williams's internalism (as understood on the reasons^{H} interpretation) true? I think that it is, and defending that claim will be the subject of Chapter 4.

Chapter 4: A Defense of Internalism on the Reasons^{H} Interpretation

In the previous three chapters I both explained what I think is the correct interpretation of internalism—the reasons^{H} interpretation—and built a strong case that it is the correct interpretation. However, even if the reasons^{H} interpretation of internalism is the correct interpretation, we are still left with the question of whether the position itself is true. Many philosophers are skeptical with respect to the truth of internalism. However, I think that most, if not all, of those who have rejected internalism have done so because they misunderstood the actual nature of the position. Most philosophers have taken internalism to be concerned with reasons^{E}. The truth of internalism on *that* interpretation—the reasons^{E} interpretation—*is* very questionable. But, presuming that I am correct that internalism is actually concerned only with reasons^{H}, then the truth of internalism is much less questionable. In fact, I think it is almost indubitable (at least as much as any position in philosophy can be).

Because there are now two different general understandings of internalism, it will help to coin some new terminology. The internalism which I am concerned to defend in this chapter rests on the reasons^{H} interpretation. I will label it internalism^{H}. Most readers of Williams however, have taken internalism to be about reasons^{E}. I will refer to that interpretation as internalism^{E}. Correspondingly, I will refer to externalism about reasons^{E} as externalism^{E}, and externalism about reasons^{H} as externalism^{H}.

So, my aim in this chapter is to defend the truth of internalism^{H}. The defense will come in three parts. First, I will provide a positive argument in favor of internalism^{H}—in particular I will argue both that we should accept the distinction it makes between reasons^{E} and reasons^{H}, and,

most importantly, that we should accept its claim that an agent *has* a reason to φ only if the agent is capable of being motivated to φ for that reason.

The second part of the defense will be the lengthiest and most important part of the defense of the truth of internalism^{H}. In this section I will respond to five objections to internalism. Bernard Williams's internalism has not suffered from a lack of detractors. However, if the reasons^{H} interpretation of Williams is accurate, then most, if not all, of the criticisms of the detractors are misguided. Therefore, the general line of my response to these objections will be that they are not relevant to internalism^{H}. In particular what I will show is that, because those who have raised the objections have interpreted internalism as internalism^{E}, none of the arguments against internalism set forth by the objectors are cogent or sound if we take them to be addressing internalism^{H}. An important implication of this is that internalism^{H} is compatible with externalism^{E}. That is, those who believe that there can be reasons^{E} which are unrelated to an agent's motivations can consistently accept that *reasons^{H}* must be related to an agent's S. Externalists about reasons^{E} can, and should, accept internalism about reasons^{H}.

In the third and last part of the defense of the truth of internalism^{H} we will consider one further objection, one which is not an objection to the substance of internalism, but instead to the formulation of it as given by Williams. I will explain why the objection rests upon a misunderstanding of the formulation.

The upshot of this chapter will then be that there is strong reason to accept both internalism^{H} and Williams's formulation of it. That is, we have sufficient reason to believe that A has a reason to φ only if there is a sound deliberative route from A's subjective motivational set to A's φ-ing; or, as I have put it, that an agent has a reason to φ only if it is within the agent's capacity to be motivated to φ.

I. Why we should think internalism^{H} is true

So, why should we think that internalism^{H} is true? The main reason is that it is true by mere definition—in particular, the definition of what it is to be a reason^{H}. In Chapter 1 we defined reasons^{H} such that for an agent to have a reason to φ, the agent must be capable of φ-ing for that reason. So, for example, if an agent is physically incapable of being motivated to φ for reason *r*, then *r* cannot be a reason that the agent *has* for φ-ing. Likewise, it would seem to follow that if an agent is incapable of being motivated to φ, then the agent does not have a reason to φ. And that claim just is the claim of internalism^{H}.

Remember that in Chapter 1 it was pointed out that, in the context of internalism, having a reason to φ is not the same as being in "possession" of a reason to φ (whatever it might mean to possess a reason to φ). We saw through an analogy with reasons to believe that it appears possible for an agent to be in possession of a reason and yet not be capable of believing in accordance with the reason. I may have seen someone who looks like a friend of mine commit a crime, but because of my other beliefs about the friend, in particular about his character, I may not be capable of believing he committed the crime. In this case, since I saw someone who at least looked like my friend commit a crime, it seems as though I am in possession of a reason to believe that my friend committed a crime, despite my inability to actually believe that he did. Likewise, for reasons to act it might be possible for an agent to be in possession of a reason to act (again, whatever that might mean) and yet not be capable of acting upon the reason.

Therefore, since "having a reason" is not the same as possessing a reason, it may be possible for an agent to have *a motive* (not a reason) related to some reason, and yet still it be the case that they do not have a reason to φ on the basis of that reason. And that is because the

motive could be heavily outweighed by other motivations of the agent such that it is not within the agent's capacity to act on the reason. For example, I may have a little motivation to exercise, but my motivation to sit at home and rest and watch TV may be so strong that I am actually incapable of exercising. Hence, despite my motivation to exercise, I still do not *have* a reason to exercise. Since that is not something I am actually capable of doing. By definition, to have a reason to φ requires that an agent be capable of φ-ing for that reason; and internalism merely specifies that motivational limitations constrain what reasons an agent is capable of acting upon, and so constrain what reasons an agent has.

Given that internalism^{H} appears to be true by definition, the burden is really upon the externalist^{H} to show that internalism^{H} is false. The question that externalists about reasons^{H} must answer is this: if other incapacities (physical, etc.) to perform an action for a reason *r* preclude *r* from being one that the agent has, why does a motivational incapacity not have the same effect? It is difficult to see what they could point to which would show that motivational incapacities are excluded from consideration in determining what actions an agent has reason for performing.

This is especially so if we consider the following scenario. Imagine that Allie is at the beach but has no intention to swim. When she was in high school she worked as a lifeguard at the ocean, and she rescued a child from a riptide, but almost drowned in the process. As a result, she has never had a desire to get back in the ocean. But she still goes to the beach in order to soak up the sun, feel the sand underneath her feet, and to enjoy the commotion of other beach-goers. Today, while reading a book from her chair, Allie notices that a teenager (who is in the ocean by himself) seems to be struggling to keep his head above the water. There are no lifeguards at this beach. So, what should she do? Since she has the physical capacity to swim and has the training to do so, it seems that there is most reason^{E} for her to run into the ocean and save

the teenager. However, due to her previous terrifying experience, Allie has such an aversion to going into the water that she could never bring herself to go into the water. For Allie, it is not within her capacity to go into the ocean to perform the rescue. What externalists about *reasons*^{E} have wanted to say is that, even if she does not have the motivation to perform the rescue, there is still a reason^{E} (in fact *most* reason^{E}) for her to do so. For various reasons they have taken Williams's internalism to deny that this is the case—since they have interpreted it as internalism^{E}. On the reasons^{H} interpretation, however, internalism does *not* deny that there is a reason^{E}, or even most reason^{E} to φ (though it does not affirm it either, since it is not a positive account of what generates reasons for action). It only denies that that Allie *has* a reason to φ.

Let us assume that the externalists^{E} are correct that there is most reason^{E} to perform the action (a position with which I wholeheartedly agree). Even though there is most reason^{E} to go into the ocean to perform the rescue, and that, were Allie fully practically rational she would do so, there is the question of what Allie, *as she is*, should *actually* do. Given her incapacity, she cannot actually go into the ocean to perform the rescue herself (at least of her own volition). We can then ask, of the actions which she *is* capable of performing, which one is there most reason to perform? That of course is just the question, what does Allie *have* most reason to do?

The point of this scenario is to show that even if we can fault an agent for not being capable of performing the action which there is most reason^{E} to perform, that still leaves us with the question of what action—which the agent is capable of performing—the agent should perform. And it is this latter issue which is the concern of Williams's internalism. He is claiming that if an agent is incapable of being motivated to φ, the agent cannot have most reason to φ—even if there is most reason^{E} to φ. What the agent should actually do is constrained by the motivational capacity of the agent. The only reason to deny Williams's internalism is if you think

that an agent should actually φ even though he is not capable of performing the action. But that position is self-contradictory.

That the externalist position is self-contradictory might raise a worry about the accuracy of my interpretation. It might seem that Williams would not have attacked what is a blatantly self-contradictory position; nor would he have needed to devote so much writing (multiple articles over twenty-plus years) to show the externalist position to be false. However, the fact that the externalist position on the reasons^{H} interpretation is self-contradictory actually supports my interpretation. Consider the following claims by Williams. In IER he writes, "The sorts of considerations offered here strongly suggest to me that external reason statements, when definitely isolated as such, are false, or *incoherent*, or really something else misleadingly expressed"[1] (emphasis mine). Or, in IROB he states, "I do not believe, then, that the sense of external reason statements is *in the least clear*, and I very much doubt that any of them are true"[2] (emphasis mine). Given these statements by Williams which express his doubt about the coherence and clarity of external reason statements, it is not surprising that externalism^{H} turns out to be self-contradictory.

II. Responses to objections to internalism

Given that externalism^{H} seems to be a self-contradictory position, it seems that the truth of internalism^{H} is largely indubitable. However, I think there might be some lingering doubts by externalists^{E} about the truth of internalism^{H}. Therefore, in order to dispel those doubts, in this section I will respond to the most prominent objections to internalism. The objections that we

[1] Williams, "Internal", 111.
[2] Williams, "Internal Reasons", 40.

will consider were raised in response to the two standard interpretations of Williams, ones which understood his internalism to be concerned with reasons^{E}, and not reasons^{H}. Therefore, they were not directly raised against internalism^{H}.

The general theme of my responses to these objections will be that they are unsuccessful because they have misinterpreted Williams's internalism as being about reasons^{E} and not reasons^{H}. In this section we will look at five different objections to internalism. To give some idea of where we are headed, let me briefly explain each of the objections we will encounter. The first is not an objection to internalism itself, but instead to Williams's claim that the interrelationship principle gives us reason to accept internalism. The objection, given by David Sobel, claims that the interrelationship principle is either too weak to support internalism, or if we try to make it strong enough to support internalism, it becomes identical with internalism, resulting in a circular argument. So, even if Williams's internalism is true, Williams has not provided us with an argument which shows that it is. I will argue that Sobel's objection rests on a misunderstanding of both the interrelationship principle and internalism.

In contrast to the first objection, the second through fifth objections are objections to the internalist position itself. The second objection claims that internalism presupposes an instrumental theory of practical reason. Although in Chapter 2 I already showed that the truth of *internalism^{H}* does not depend on any particular theory of practical reason, it will be beneficial to respond to this objection in order to see *why* the instrumentalist objection to internalism does not hold for internalism^{H}.

The third objection to internalism claims that internalism unjustifiably denies a volitionalist account of practical *agency*. This objection is distinct from the previous one. This objection claims that internalism presupposes that it is not possible for an agent to acquire a

motivation to perform an action which is not related to his existing motivations. The claim is *not* that internalism presupposes that an agent can only deliberate about his motivations. Rather, even if internalism does allow that an agent *can* deliberate about considerations other than his motivations, the objection claims that internalism (at least implicitly) denies that the agent can acquire the motivation to act on one of those considerations if it is not already related to his current motivations. My response will be that this objection has too narrow an understanding of what motivational elements internalism allows to be included in an agent's motivational set.

The fourth objection we will consider takes internalism to deny that there can be a reason^{E} to φ if an agent is incapable of being motivated to φ. Even if a high school student cannot be motivated to study for a test which will significantly affect his chances of getting into a good college, there is still a reason^{E} for him to study. As is likely clear by now, I will argue that internalism^{H} does *not* deny that there is a reason^{E} for the agent to study. Internalism^{H} is only a claim about reasons^{H} and not reasons^{E}, and so it would only deny that the student *has* a reason to study. It would not deny that there is a reason^{E}.

The fifth and final objection we will consider in this section is that internalism erroneously denies the existence of some moral reasons and/or responsibility. The general gist of the objection is that internalism implies that if an agent is not motivated to act on a putative moral reason^{E}, then there is not a moral reason^{E}, nor is the agent morally responsible for not acting on the putative reason^{E}. For example, if an agent is unmotivated to help alert someone of an oncoming danger, then there is not a moral reason^{E} for him to do so, nor is he responsible for not alerting the other person (and all of the consequences which result from his inaction). In response, I will point out that internalism^{H} does not make any claims about moral reasons or

moral responsibility, and so whether the agent would be responsible depends instead on the account of moral reasons and responsibility.

As I said, the main tenor of my response to the above objections is that they are unsuccessful because they misinterpret Williams's internalism as being concerned with reasons^{E}. But, because I am arguing that they are unsuccessful due to misinterpreting Williams, it might appear that I am distorting the arguments of the philosophers who gave them. The objectors might (reasonably) claim that their arguments should only be assessed on their success in demonstrating internalism^{E} to be false (or at least dubitable), and not internalism^{H}. However, my aim is mainly to clear up any remaining confusion over the nature of the internalist^{H} position, and so the point of considering these objections is to show that the concerns about *internalism^{E}*—which prompted the objections—are not concerns for internalism^{H}. The upshot of my responses to these objections will be that there is no good reason that an externalist about reasons^{E} should not be an internalist about reasons^{H}.

1. Objection: The interrelationship principle cannot support internalism

This first objection is not a direct objection to Williams's internalism, but instead it is an objection to his claim that the interrelationship between normative and explanatory reasons provides support for internalism. We can capture that claimed interrelationship with the following (roughly stated) principle: in order for a consideration to be a normative reason for an agent to act, it must be possible that the consideration also serve as part of an explanation of the agent's performing the action.[3] According to David Sobel, either the interrelationship principle

[3] I have avoided using the interrelationship principle that was formulated as IP in Chapter 2 because it would presume against one of the interpretations of that principle which Sobel considers (which we will see below).

(what Sobel calls the "explanation condition") is too weak to support internalism, or, if it is modified in order to be strong enough to support the internalist thesis, it becomes equivalent to the internalist thesis.[4] The latter is problematic because the argument would be circular. In the following, we will take a look at why Sobel claims that the interrelationship principle faces this dilemma, and I will then explain why his argument does not apply to internalism on the reasons^{H} interpretation.

To explain the problem with the interrelationship principle**,** Sobel quotes the following statement by Williams.

> [A fundamental motivation of the internalist account] is the interrelation of explanatory and normative reasons. It must be a mistake simply to separate explanatory and normative reasons. If it is true that A has a reason to φ, then it must be possible that he should φ for that reason; and if he does φ for that reason, then that reason will be the explanation of his acting.[5]

As we saw in preceding chapters, the meanings of "normative" and "possible" are ambiguous. Like most others, Sobel takes Williams to be using "normative" in the agent-neutral sense. So he takes internalism to be concerned with reasons^{E}. He then goes on to consider what notion of possibility Williams has in mind. Sobel begins by focusing on Williams's claim that that if an agent has a reason to φ, it must be possible that the agent would φ for that reason. Sobel considers three potential interpretations of that claim, all with an eye towards seeing how they might support the internalist claim that "A has a reason to φ only if there is a sound deliberative

[4] Sobel, "Explanation", 218.
[5] Williams, "Internal Reasons", 38-9. Quoted in Sobel, "Explanation", 219-220.

route from A's subjective motivational set to A's φ-ing".[6] Not surprisingly, the three different interpretations all vary with respect to their interpretation of "possible". To understand Sobel's argument, it is important to note that when the internalist thesis requires that there be a sound deliberative route from A's motivational set, Sobel interprets that to mean that there is a sound deliberative route from the motivational set of the actual agent A, not some idealized version of A.

The first potential interpretation, what Sobel calls *Explanation I*, takes the above claim to mean that "it is a necessary condition of A having a reason to φ that there be a possible world in which A φ's."[7] Sobel argues that if this is the correct interpretation, it is too weak to support internalism. That is because even if there is a possible world in which A φs, that does not entail that there is a sound deliberative route from the actual agent's S to his φ-ing for reason *r*. This is due in large part because the possible world in which A φs does not even have to be one where he φs *for* the putative reason. Suppose that we are considering whether there is a reason^{E} for an agent to study for a test for the reason *T* that it will improve his grade. If a high school student has absolutely no interest in studying to improve his grade, he may be motivated to study if he knows that it will impress a girl he is interested in. So, there is a possible world in which A φs—the one in which he studies to impress the girl. But, given that the student has no interest in studying in order to improve his grade, we know that there is not a sound deliberative route from the student's S to his φ-ing for reason *T*. In this case the *Explanation I* condition is met, but the internalist condition is not. Hence, *Explanation I* does not provide sufficient support for the internalist thesis. Sobel recognizes that part of the inability of *Explanation I* to adequately

[6] Williams, "Postscript", 91
[7] Sobel, "Explanation", 220.

support the internalist thesis is that it does not require that, in the possible world wherein the agent φs, the agent φs for the reason (what Sobel calls the "consideration") in question.

Given the problem with *Explanation I,* Sobel offers a second potential interpretation of Williams's explanation condition, what Sobel calls *Explanation II*. According to it, "if consideration C gives A a reason to φ, it must be the case that A can φ *and* that in some possible world in which A does φ, his doing so is explained by his being motivated by C".[8] This interpretation at least remedies the glaring problem with *Explanation I.* However, *Explanation II* also does not adequately support the internalist thesis. As Sobel puts it, "Explanation II is insensitive to the distinction between A's being motivated by C via a sound deliberative route and A's being so motivated in other ways (such as radical brain surgery)."[9] For example, even if there is a possible world in which the student studies *in order to improve his grade* (for consideration C), that does not entail that there is a sound deliberative route from the student's subjective motivational set (as he is) to studying for C. It could be that a mad scientist has rewired his brain so that he is motivated by C to study. So, the problem with *Explanation II* is that it does not specify *how* the student comes to be motivated by C to φ. Any old means will do. But that cannot support the internalist claim that there must be a sound deliberative route from the agent's S to φ-ing for C. So perhaps that is not Williams's understanding of the interrelationship principle.

Given the problems with *Explanation I and Explanation II,* Sobel concludes that Williams must have understood the explanation condition to have a further requirement. He must have meant for it to require that the possible world in which the agent φs for C be one which he is able to reach through sound deliberation from his present world. So, Sobel says the

[8] Sobel, Ibid., 222.
[9] Ibid.

explanation condition might be *Explanation III*: " a jointly necessary condition of consideration C providing A a reason to φ is that (1) A could φ; (2) in some possible world in which A φs, his φ-ing can be explained by means of his contemplation of, and subsequent motivation by, C; and (3) in some possible world in which (1) and (2) are the case, A is deliberating soundly from his actual subjective motivational set."[10] This does ensure that the possible world where A φs for reason C is one in which he does so as a result of sound deliberation. However, the problem with this interpretation is that it just *is* the internalist thesis. Therefore, it cannot serve as the basis for the internalist thesis.

If Sobel is correct, this does not show that internalism is false. But it does mean that Williams has not provided independent support for the truth of internalism. He has not given those who presently reject internalism a genuine reason to accept it.

However, Sobel's argument is flawed. Sobel's argument that the interrelationship principle cannot properly support the internalist thesis rests upon a faulty interpretation of Williams. I think Sobel is correct that if we interpret internalism as being concerned with reasons^{E}, then the interrelationship principle cannot support internalism.[11] But, as I have argued, "possibility" is to be interpreted as "within an agent's capacity" and "normative reason" as "reason^{H}".[12] As was already shown in Chapter 2 the interrelationship principle *is* able properly to support internalism on the reasons^{H} interpretation. But, given Sobel's worry that it cannot, I want to briefly explain why the reasons^{H} interpretation avoids the problem which Sobel thinks he has identified.

[10] Ibid., 223.
[11] I think this is especially the case because the interrelationship principle would very likely be false if it was a claim about reasons^{E}.
[12] As well, Sobel did not even consider the other meanings of "possibility" that we considered in Chapter 3. But since I have already shown why they are faulty, I will not consider them here.

Internalism on my interpretation is concerned with the reasons that an agent is actually capable of acting upon. So, on the reasons^{H} interpretation the interrelationship principle merely claims that, in order for an agent to have a reason to φ, it must be within the agent's capacity to φ. That is a general principle, one which does not apply specifically to an agent's motivational capacities. Given that having a reason requires being capable of acting upon the reason, the truth of the interrelationship principle seems unquestionable. And, since it is a general principle, it can be applied to particular limitations on an agent's capacity to perform an action. For example, given that it is true, it follows that if an agent is not physically capable of performing an action, then the action is not one that the agent has a reason for performing. And, likewise, it should also follow that an agent who is motivationally incapable—i.e., psychologically incapable—of performing some action does not have a reason to perform the action.

At this point we might notice that there are two components of the internalist formulation which the interrelationship principle does not include. It does not say anything about 1) there being a sound deliberative route, or 2) that the agent must act *for* the reason in question. Since it does not, we might wonder how it can support the internalist claim that there must be a sound deliberative route from an agent's S to φ-ing for that reason? It can because those two components of IP merely follow from consideration about what it is for an agent to be motivationally capable of performing an action (both were argued for in Chapter 2).

The first component, the inclusion of the sound deliberative route, results from the need to get an accurate understanding of the agent's actual motivations. The agent who appears to be motivated to mix the petrol with tonic and drink it, does not truly have that motivation. He is motivated to mix what is in the glass with tonic only because he *thinks* gin is in the glass. So, we should not say that he has the motivation to mix petrol with tonic and drink it. The sound

deliberative route filters out such motivations—ones which are dependent upon a false belief—from the agent's motivational set.[13]

The second component, the inclusion of "for that reason"—follows from the consideration that agents can φ for different reasons, and the motivation to φ for reason r^1 is not the same as the motivation to φ for reason r^2. If a particular reason is one that an agent has to φ, the agent must be motivationally capable of φ-ing for *that* reason, and not some other reason which supports φ-ing. For example, suppose someone has the motivation to look good, but does not care about being healthy. If so, he could be motivated to exercise (φ) for the sake of looking good (r^1), but not for the sake of being healthy (r^2). Hence, despite the ability of the agent to be motivated to φ, r^2 (exercising improves health) is not a reason he has for φ-ing, because he is not capable of being motivated by that reason to φ.

So, on the reasons^H interpretation, the interrelationship principle can support the internalist thesis. That is because IP is merely a general principle about what is required for an agent to have a reason to perform an action: It must be within an agent's capacity to perform the action. And the internalist thesis is merely the result of the application of IP to an agent's motivational capacity to perform an action. It only fills out in fuller detail what it is to be motivationally capable of performing an action.

This objection by Sobel was only about the ability of the interrelationship principle to support internalism, and not an objection to the internalist position itself. Next, however, we will consider three objections to the position of internalism.

2. Objection: Internalism presupposes a quasi-instrumental theory of practical reasoning

[13] Keep in mind that internalism is not concerned with what it is epistemically rational for an agent to do. That is, it is not concerned with what the agent should do, given his beliefs. So, we do not want to include actions which an agent is capable of being motivated to perform on the basis of false beliefs. I argued for this in Chapter 2.

Since internalism requires that an agent's reasons be related in some way to his or her motivational set, one of the common objections to internalism is that it presupposes an instrumental or perhaps quasi-instrumental theory of practical reason. We saw this objection first come up in Chapter 1. And, in Chapter 2 I showed that Williams's argument against external reasons did not presuppose any theory of practical reason on the reasons^{H} interpretation. What I want to do now is to show, by considering a specific objection to the effect that internalism presupposes instrumentalism, the misunderstandings that have often led people to conclude that it does. We will take Brad Hooker's objection as representative of this type of objection. What can be said in response to it largely holds for other instrumentalist objections as well.

Hooker claims "[t]he dispute between Williams and the external reasons theorist is ultimately over the starting points of practical deliberation."[14] That is, whether internalism is true depends upon what type of practical deliberation is legitimate. According to Hooker, internalism requires that an agent, in deliberating about what there is reason for him to do, must deliberate only about the elements in his motivational set. Hooker does not say whether he thinks there are legitimate forms of practical deliberation that do not start with the agent's motivations. He is not arguing that internalism is false. Instead, he is only arguing that the truth of internalism depends upon the claim that all practical deliberation must start from an agent's motivations.

Before we consider why Hooker claims that internalism depends upon, in my terminology, a quasi-instrumental theory of practical reason, it is important to note that he does recognize that Williams has a wide conception of an agent's motivational set—one which is composed of more than just the agent's desires. Hooker acknowledges Williams's allowance that an agent's S can also include the agent's "dispositions of evaluation, patterns of emotional

[14] Hooker, "Williams' Argument, 42.

reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent".[15] Though this is not a narrow instrumentalism, Hooker contends that, nonetheless, it is still instrumentalist in nature (what I call "quasi-instrumentalist").

To provide support for his claim, Hooker begins by reconstructing William's argument against external reasons (which is different from my formulation of it). The argument is as follows.

> 1. Given that one had made oneself aware of the relevant empirical facts, including (perhaps through imagination) facts about what some alternative outcome would be like, one would be engaging in rational practical deliberation if one were (a) ascertaining what way of satisfying some element in one's subjective motivational set would be best in the light of the other elements in the set, (b) deciding which among conflicting elements in one's subjective motivational set one attaches most weight to, or (c) 'finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment'. (This account of rational practical deliberation can be distilled from pp. 103-5.) Thus rational practical deliberation in each of its forms has as its starting point the subjective motivational set had by the agent prior to this deliberation.
>
> 2. A reason is *internal* just if it can be reached by rational practical deliberation which starts from the agent's antecedent subjective motivational set.
>
> 3. From (1) and (2) it follows that any reasons the agent arrives at by *rational practical deliberation* will be internal reasons. (That is, if I could, by means of deliberating in any

[15] Williams, "Internal", 105.

of the ways mentioned in (1), arrive at a new belief about what there is reason for me to do, then there was an internal reason for me to do the act in question to begin with (p. 109).)

4. What there is reason for me to do is determined by what I would, if I deliberated rationally (and knew all relevant empirical facts), find there is reason for me to do.

5. From (3) and (4) it follows that there are no such things as external reasons.[16]

Hooker claims that whether this argument is sound depends upon whether the only type of practical reasoning is instrumental reasoning. He notes that in premise (1) all of the examples of practical deliberation—which are extracted from Williams's list of examples from page 104 of IER—are instances of deliberation *about* the agent's motivations. Hooker takes the list to be exhaustive of the types of practical deliberation which Williams allows. Therefore, assuming (1) is true, then it would be the case that any reason an agent has to act on can be reached through practical deliberation from the agent's motivations. And, since in premise (2) Williams defines an internal reason as one which can be reached through rational practical deliberation which starts from the agent's antecedent motivations, it turns out that all reasons are internal.

According to Hooker, the problem with this argument is that those who think that rational practical deliberation can be about something other than one's own motivations will not accept premise (1). Practical deliberation, it can at least be argued, might start from considerations external to an agent's motivations. To reach the conclusion to eat lunch, it is not as though I can only come to that conclusion on the basis of the fact that I have the motivation to eat. Instead, it

[16] Hooker, "Williams's Argument", 42-3.

could be that I deliberate and realize that there is a reason to eat lunch (because it would remove my hunger pains, or because it would provide nutrition for my body, etc.), and therefore there is a reason to eat lunch, irrespective of whether I have the motivation to do so. If this is an instance of a genuine type of practical deliberation, the reason for eating lunch would be an external reason. Though this is just one example, it illustrates Hooker's point that whether internalism is true depends upon our conception of rational practical deliberation. If there are types of practical deliberation which are not merely about the elements of an agent's S, then premise (1) would be false. And if premise (1) is false, then there could be reasons for action which can be reached through practical deliberation which does not start from the agent's motivations. That is, there could be external reasons.

The problem with Hooker's construal of Williams's argument is that it misrepresents Williams's position. He makes it appear as though Williams only allows agents to reason about how to satisfy their desires, and so only allows for instrumental or quasi-instrumental accounts of practical reasoning. It is true that in Williams's list of types of practical reasoning on page 104 of IER, he only lists instrumental types. Since I have already argued in Chapter 1 that Williams's internalism should not be taken as relying upon an instrumental account of practical reason, I will not provide an in-depth defense of that claim here. But I do want to address the likely source of confusion for Hooker (and others).

Williams's list does seem to strongly suggest some type of instrumental theory of practical reason. However, the list comes prior to his claim on page 105, wherein he states, "I have discussed *S* primarily in terms of desires, and this term can be used, formally, for all elements in *S*. But this terminology may make one forget that *S* can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various

projects, as they are abstractly called, embodying commitments of the agent."[17] So, perhaps when Williams gave the list on page 104, he was doing so from the "desire-only" perspective. And, once he admits of other types of elements in an agent's S, that opens up the possibility of additional types of practical reasoning. In other words, were the only elements of an agent's S to be the agent's desires, then there could only be instrumental reasoning; but, since an agent can be motivated by elements other than just his or her desires, non-instrumental reasoning is also possible. Given Williams's repeated assertion that internalism does not rely upon an instrumentalist theory of practical reasoning, we have good reason to take this latter interpretation to be correct.

With the rejection of the instrumental-only interpretation, premise (1) should be amended to *allow* for all possible types of practical reasoning. (Therefore, we will just remove the list of types of deliberation as well as the last sentence that "Thus rational practical deliberation in each of its forms has as its starting point the subjective motivational set had by the agent prior to this deliberation" is not a part of the premise.) Doing so has a significant impact on the soundness of the argument as Hooker has formulated it. Whether the argument is successful will now depend on whether we interpret "reason" in the argument as reason^{E} or reason^{H}. This is best exemplified in the fact that the truth of premise (3) is now questionable. Whether it is true depends upon which interpretation we take.

Consider the argument if we interpret "reason" as reason^{E}. When (1) only allowed for instrumental reasoning, it did follow from (1) and (2) that all reasons^{E} are internal. And that was because (1) only allowed for rational practical deliberation that starts from an agent's motivations and (2) defined an internal reason just as one which can be reached through deliberation that starts from an agent's motivations. But, if (1) allows for deliberation which does

[17] Williams, "Internal", 105.

not start from an agent's motivations, then (3) does not follow. It is not necessarily the case that any reasons the agent arrives at by *rational practical deliberation* will be internal reasons. There may be reasons^{E} which are external to the agent's motivations. We see here that Hooker is correct that Williams's argument against external reasons—if interpreted as reasons^{E} (as Hooker and others interpret it)—is sound *only if* rational practical deliberation only starts from an agent's motivations.

But, if "reason" is interpreted as reason^{H}, the implications for the argument are substantially different. First, like we did with our consideration of Williams's argument on the reasons^{E} interpretation, we amend (1) to allow for all types of practical reasoning. Second, throughout the argument we should understand all of the references to "reason" or "reasons" as referring to reasons^{H}. We can now evaluate the soundness of the argument on the reasons^{H} interpretation.

Despite the allowance for non-instrumental types of reasoning, the claim in (3) that all reasons^{H} are internal is still true. However, unlike the original argument, the truth of (3) does not follow from (1) and (2). It follows merely from (2) and its definition of an internal reason^{H}. There is a reason^{H} for an agent to act only if the agent is capable of acting upon the putative reason. If there is a *putative* reason^{H} which cannot be reached from the agent's motivations—that is, if it is not within the agent's capacity to be motivated by the reason through deliberation—then it is not a reason^{H}. So, premise (3) is correct that all reasons^{H} are internal (though it does not follow from premise (1)). And if *all* reasons^{H} are internal, then it follows that premise (5) is true—no reasons^{H} are external. And this is so even though premise (1) allows for all possible types of practical reasoning. It is true even if it allows for types of practical reasoning beyond instrumental or quasi-instrumental reasoning.

So, we have seen that Hooker's objection to Williams's argument—that it relies upon an instrumental theory of practical reasoning—follows only if we understand it to be about reasons^{E}. And, since I have established in the previous three chapters that it is best understood as being about reasons^{H}, Hooker's objection is not relevant to internalism^{H}. The truth of internalism^{H} does not depend upon an instrumental theory of practical reason.

3. Objection: Internalism unjustifiably denies a volitionalist account of practical agency

Despite the fact that Williams's internalism allows for non-instrumental practical reasoning, Jay Wallace objects to its conception of practical agency. He claims that internalism unjustifiably restricts what motivations an agent is capable of acquiring to only those motivations which have some relationship to the agent's existing motivations. Therefore, Williams's internalism (at least implicitly) denies *volitionalism* about practical agency—the idea that "there is an important class of motivational states that are directly subject to our immediate control."[18] According to volitionalism, there are motivations which we can acquire merely through the use of our rationality and which are in no way dependent upon our existing motivations.[19] Internalism presupposes an opposing view of practical agency, *hydraulicism*. Hydraulicism is a view about human motivational psychology which claims that what we are capable of being motivated by (and therefore what actions we are capable of performing) is tied to our present motivations.[20, 21]

[18] Wallace, ibid., 236.
[19] Shafer-Landau calls this ex nihilo motivation. It is also similar to the type of motivation Thomas Nagel defended in *The Possibility of Altruism* (Oxford: Clarendon Press, 1970). See especially page 111.
[20] Wallace, "Three Conceptions of Rational Agency", *Ethical Theory and Moral Practice* 2 (1999): 233.
[21] Wallace rejects hydraulicism in part because he takes it to entail that actions and practical inferences are merely the result of psychological forces to which we as agents are passive.

Wallace's objection is especially problematic for internalism because he claims that *if* internalism allowed for a volitionalist conception of practical agency (which he does not think it can), it would then cease to be an "interesting" position. It would not be interesting because an agent's present motivations would as a result not place any actual restrictions on the normative reasons of the agent. Since it would be possible for the agent to acquire a motivation entirely unrelated to his existing motivations, his current motivations do not constrain what motivations the agent is capable of acquiring, and therefore also does not constrain what actions he is capable of performing.

If Wallace is right, internalism faces a dilemma. If it denies volitionalism, then internalism is unjustified (as a strong argument against volitionalism has not been provided), or if it allows for volitionalism, then it is not an interesting thesis. I will argue that this is not a genuine dilemma, as internalism does allow for volitionalism, but yet it is still an interesting thesis about agents' reasons for action. That is, it does provide a restriction on an agent's normative reasons.

Wallace succinctly explains his objection in "Three Conceptions of Practical Agency". He writes:

> Internal reasons presuppose that some disposition to action is already to hand, and reflection on such reasons can therefore generate a new motive by tapping into the agent's preexisting motivations (so to speak). But matters are not so simple with external reasons. These reasons precisely do not presuppose that an appropriate source of motivation is already present. So it must be explained how reflection in terms of external reasons could generate by itself a new motivation to action. Internalists such as Williams

> claim that this burden of argument cannot be met. It is not entirely clear, however, what the force of this challenge is supposed to be. One is initially tempted to respond on behalf of the externalist with a reminder that there are no a priori constraints on causal relations. If the question is, how could reflection by itself produce a new motivation to action?, the answer might simply be that there is no general reason to suppose that it could not. That is, in advance of empirical inquiry we have no more reason to exclude the possibility that externalist reflection might generate a new motive than we have to rule out any other kind of causal consequence.[22]

The nature of this objection can perhaps best be seen in the following example. Suppose that Andrew sees a child riding her bike down the street, unknowingly headed towards a giant pothole which, were she to run into it, would cause her to crash. Suppose that we think that Andrew should alert the child to the pothole. What must be true about Andrew's S if he is going to be able to acquire the motivation to alert the child? Volitionalism claims that there can be no elements in his S and yet still he could acquire the motivation to alert the child. That is because there can be objective features of this scenario which generate reasons^{E} to act which Andrew can respond to, given that he is rational. Purely in virtue of his rationality, he can acquire the motivation to perform the action which there is most reason to perform. Wallace thinks that internalism rejects that idea. Instead, internalism claims that whether Andrew is capable of coming to have the motivation to alert the child depends upon his pre-existing motivations. If he has a desire to prevent others from experiencing pain, or if he is motivated to perform actions which would make him appear heroic, and so on, then internalism would allow that he could come to have the motivation. But, if he has no motivations which are properly related to alerting

[22] Ibid., 220-1.

the child, then internalism denies that he could come to have the motivation to do so. Wallace objects and asks, where is internalism's argument for such a restriction? If objective features of the situation generate reasons to alert the child, then it is at least possible (absent arguments to the contrary) that Andrew, as a practically rational agent, could come to have the motivation to alert the child even if he does not currently have a related motivation.

This objection is not to be equated with the objection of Korsgaard and Smith, who argued that Williams's internalism is faulty for claiming that the reasons that a fully rational version of an agent would be motivated by depends on the existing motivations of the actual agent. That is an issue related to our conception of practical reason—i.e. of what generates reasons for action. Korsgaard and Smith objected to what they took to be Williams's claim that there can exist a reason^{E} to φ only if an agent currently has a motivation related to reason^{E}. That is, the existence of a reason^{E} for action depends on an agent's existing motivations. Wallace's objection, however, is not concerned with whether the existence of a reason^{E} depends on an agent's motivations. Instead, it is concerned with the issue of our conception of practical agency, and in particular with the motivational capacities of agents. He takes Williams to be claiming that it is possible for an agent to be motivated by a reason^{E} only if the agent has an existing motivational element related to reason^{E}. According to volitionalism, it is possible for an agent to acquire a motivation entirely unrelated to his present motivational set. Therefore, Wallace thinks that internalism is unjustified in claiming that what it is within an agent's capacity to be motivated by is necessarily constrained by the agent's S.

Does internalism unjustifiably restrict what motivations it is within an agent's capacity to come to? No. It must be remembered that Williams construes the content of the agent's S very broadly, so as to include not only the agent's desires, but also his "dispositions of evaluation,

patterns of emotional reaction, personal loyalties, and various projects…embodying commitments of the agent".[23] As I mentioned before, we should think of S as composed of *whatever* it is that is capable of motivating the agent. So, if an agent is composed such that, through the mere recognition of objective reasons^{E} to φ he is capable of acquiring the motivation to φ, then that capacity is an element in his S. In other words, internalism does not deny volitional practical agency (VPA). However, for any agent who has VPA, that is a motivational capacity which is an element of the agent's S.

It might be wondered whether my claim that internalism allows for volitional practical agency is consistent with Williams's account of internalism. Certainly Williams tended to characterize the agent's S as composed of motivations (narrowly construed, e.g. desires, etc.). That, I think, is a reflection of the fact that Williams's own conception of practical reason was desire-based. But he denied that it played a role in his account of internalism.[24] Instead, he conceded that non-desire-based accounts of practical reason *could be* correct. In footnote #3 of IROB he says that he does not reject Korsgaard's claim that there *could be* requirements of pure practical reason on every fully rational agent—i.e reasons^{E} to act which do not depend on any agent's motivations and yet which every *fully* practically rational agent would be motivated to act on. He reaffirms this in "Postscript".[25] (We must keep in mind—as I defended in Chapter 3—that internalism is not concerned only with the reasons of a fully practically rational agent. An agent's lack of rationality can preclude him from having a reason to act which a fully practically rational agent would have.) And—this is the important part—Williams then adds that, if such

[23] Williams, "Internal", 105.
[24] On page 35 of "Internal Reasons" he claims that he thinks an agent's having a desire provides a sufficient condition for the agent to act, but he clarifies that internalism is only making the claim that it is a necessary condition.
[25] Williams, "Postscript", footnote 4, 97.

reasons exist (which he doubts[26]), then the rational agent's capacity to be motivated by such reasons would be one of the elements of the agent's S.[27] And that is precisely the position for which I have been arguing. In other words, Williams allows for the possibility that an agent can be motivated by a reason^{E} which is unrelated to her motivations (narrowly-construed,), and if so, then that capacity is a part of the agent's motivations (broadly-construed).

However, Wallace has a worry about the ability of internalism to allow for volitional practical agency. He thinks that such a capacity would likely be subsumed under the category of what Williams has called "dispositions of evaluation"; and, Wallace claims, it is misleading to construe rational motivation in terms of having a standing disposition to perform actions of a certain kind. To do so implies that when an agent responds to an objective practical reason he does so only because he already has a motivational state (narrowly construed) related to the reason. For example, on the "disposition of evaluation" conception, when Andy sees the bike-riding child in danger and sees it as a reason to act, Andy already has some character-state that tends towards helping those in danger. On the volitional account, Andy's capacity to be motivated by the objective practical reasons is not the result of his motivational state, but his rationality. His capacity to be motivated by such a reason is not due to his merely being disposed to be motivated by such reasons. Rather, he has the capacity due to his rationality.

I agree with Wallace that it would be misleading to characterize motivation on the basis of objective practical reasons as we are discussing here as the result of a "disposition of evaluation"—if that phrase is too narrowly construed. But, as Wallace notes, we can construe "dispositions of evaluation" very broadly, so as to include even volitional acquisitions of

[26] Williams, "Postscript", 94.
[27] Williams, "Values", 111.

motivation.[28] The other option is to add "volitional acquisition" to the list of elements which can compose an agent's motivational set, rather than broadening our use of "dispositions of evaluation". In the end, I do not think it is all that important which option we go with. What is important is that Williams allows for volitional acquisitions of motivation. What label we put it under is mere semantics.

However, Wallace objects that, once we allow the inclusion of volitional acquisitions as a component of an agent's S, internalism's distinctiveness is lost. He writes:

> So interpreted, however, the claim [that a necessary condition for an agent having a reason to φ is that there be a sound deliberative route from the agent's S] fails to rule out any of the interesting options in the theory of rational motivation that one might wish to endorse. In particular, it does not rule out the idea that correct reflection on one's moral reasons might suffice to generate a corresponding motivation, even if, prior to the episode of deliberation, one lacked any identifiable desire from which the new motivation might have derived.[29]

If internalism allows that an agent can acquire motivations unrelated to their existing motivations (narrowly construed), then any objective reason for action—that is, any consideration which is required by reason for a practical agent to act upon—is one that agents are capable of acting upon. It seems that the agent's motivations (narrowly construed) have no restricting force upon the agent's capacity to be motivated to act upon objective practical reasons. And, therefore, what

[28] Wallace, "Three Conceptions", 221.
[29] Ibid.

reasons an agent *has* is in no way restricted by the agent's motivational set—since any objective practical reason is one which it is possible for the agent to acquire the motivation to act upon.

What Wallace's claim fails to recognize is that, for internalism, statements about an agent's normative reasons for action are distinctively about the agent. And so, although internalism allows for volitional practical agency, whether such a capacity is included amongst a particular agent's S depends upon whether that particular agent has the capacity or not. For example, even if there is such a thing as VPA and that there is an objective reason^{E} for Andy to alert the bike rider of the pothole, it does not automatically follow that Andy *has* a reason to alert the bike rider. He only has such a reason if he is capable of being motivated to alert the bike rider. If Andy does not have VPA, or only has it to a limited degree, and so it is not within his capacity to be motivated to alert the biker rider, then he does not have a reason to do so. And that is why internalism is still an interesting thesis even if it allows for VPA. The motivational sets of different agents almost certainly have different elements, and so what reasons different agents have varies according to their different motivational capacities. True, if all existing human agents have the exact same motivational capacities, then internalism will not result in agents having different reasons (assuming they are in the same circumstances). Williams admits as much.[30] But, so long as they do have different motivational capacities, which seems to be most likely, then what reasons they have will vary.[31]

So, we have seen that, contrary to Wallace's claim, internalism does not face the dilemma of either unjustifiably denying volitionalism or being an uninteresting thesis if it does. Practical

[30] On page 114 he writes, "[the] internalist doctrine would be pointless if everyone's values and everyone's *S*, were the same; in those circumstances, anyone's deliberation could be anyone else's, and the distinction between externalism and internalism would fade away." By it "fading away", I take it that he means the *implications* of those views would fade away for the reasons^{H} of agents. All agents in the same circumstances would have the same reasons.

[31] And, even if all existing agents had the same motivations, internalism would still entail that, for hypothetical agents who had different motivational capacities, the reasons they would have would differ.

agents who have VPA—the capacity to acquire motivations merely by recognition of practical reasons—will have that capacity as a part of their motivational sets. But for those agents who do not have such a capacity, or only have it to a limited degree, their Ss will reflect that fact. And, therefore, internalism is still "interesting", because the former agents could have some reasons to act which the latter do not.

4. Objection: A reason to φ can exist even if an agent is incapable of being motivated by the reason to φ

Although internalism does not presuppose a quasi-instrumental theory of practical reasoning, nor deny a volitional account of practical agency, some have faulted it for claiming that the existence of a reason^{E} to φ depends upon an agent being able to φ for that reason. In Chapter 1 we considered an argument that Shafer-Landau gave against internalism. He asked us to imagine a person (whom we called Debbie) who is melancholic, and so has no desire or motivation to get out of her house and develop friendships with others. Because of her present motivational state, it is not within Debbie's capacity to be motivated to leave her house in order to have a social life. But, were Debbie to do so, she would find the friendships pleasurable (even if not immediately) and it would bring her out of her melancholy. Shafer-Landau states that surely there is a reason for Debbie to develop a social life even if she lacks the motivation to do so. Given how much better her life would be were she to have friends, it seems undeniable that there is a reason^{E} for her to do so. According to Shafer-Landau, since Debbie cannot be motivated to φ, internalism claims that there is not a reason^{E} for her to φ. The form of the argument, as we saw in Chapter 1, is the following.

1. If internalism is true, then there is not a reason^{E} for Debbie to develop a social life if she is incapable of being motivated to develop one.

2. There is a reason^{E} for Debbie to develop a social life even if she is incapable of being motivated to develop one.

C. Hence, internalism is false.

Despite the seeming soundness of this argument, the problem with it is that it misunderstands the type of reasons with which internalism is concerned. As I argued in previous chapters, internalism is only a claim about reasons^{H} and not reasons^{E}. The argument above is about reasons^{E} and so therefore premise (1) is false. It is not the case that if internalism is true then there is not a reason^{E} for Debbie to φ. Even though Debbie has no motivation to φ, there can still be a reason^{E} for her to φ. But, she does not *have* a reason to φ. As I stressed in the beginning portion of this chapter, internalism is concerned with what reasons an agent should actually act on given his or her limitations. Since in this scenario Debbie is not capable of acting on the reason, then it is not one that she has. It would not do any good to advise Debbie to develop friendships if she is not capable of doing so.

This distinction also allows us to respond to another prominent objection similar to Shafer-Landau's. John McDowell argues that even if an agent does not satisfy the internalist requirement, there can still be a reason^{E} for the agent to φ.[32] He sees this as following from the broadly Aristotelian idea that, in order for an agent to perform the correct action for the right reasons, the agent needs to have not only the right information, but also the right desires. So, although an agent has the right information (and so is deliberating soundly), if he lacks the right

[32] Alasdair MacIntyre argues a similar point in his *Dependent Rational Beings*. See pages 86-7.

desires he may not have the motivation to φ for reason *r*, even though *r* is the rational thing to do. Given this view of practical reason, McDowell agrees with Williams that what agents will be motivated to do after sound deliberation will vary between agents, because their starting motivational states will often vary.[33]

However, McDowell denies that such variation after sound deliberation entails that there cannot be external reasons. He claims that there can be a reason *r* to φ despite the fact that the motivation to φ for reason *r* is inaccessible to an agent even after sound deliberation. Consider a cowardly soldier who is not capable of being motivated to defend his town when it is being pillaged by the enemy. His fear of being killed is so strong that, even if he deliberates soundly and perhaps even recognizes that there is most reason^{E} for him to φ, he nonetheless is incapable of φ-ing. In this case, the soldier's lack of motivation is due to his not being fully virtuous. A fully virtuous agent would be motivated to defend the town. Therefore, there is a reason^{E} to φ, even if the soldier is unmotivated to φ for that reason. So, although McDowell agrees with Williams that sound deliberation will not issue in all agents being similarly motivated, he disagrees that such a truth entails that there is not a reason^{E} for an agent to φ. There can still be a reason^{E} for an agent to φ for *r*, even if he is not capable of reaching the motivation to φ for *r* through sound deliberation.

As should be readily apparent now, McDowell, like Shafer-Landau, has misunderstood the nature of the internalist position. Internalism is concerned with reasons^{H}, not reasons^{E}. Internalism denies that the cowardly soldier *has* a reason to defend the town. But it does not deny that there is a reason^{E} to defend it. Defending the town may be a really good action, the virtuous action even, but it is not one that the soldier *has* reason to perform given his motivational limitation.

[33] John McDowell. *Mind, Value, & Reality* (Cambridge: Harvard University Press, 1998), 95-101.

5. Objection: Internalism erroneously denies the existence of some moral reasons/responsibility

Despite the ability of internalism^{H} to accommodate external reasons^{E}, some may still think it is unsatisfactory. It might be argued that there are some reasons that an agent has no matter whether he is capable of being motivated to act on them or not—in particular, moral reasons. Even if (from the earlier example) Andrew cannot come to have the motivation to alert the child to the danger of the pothole, he still has a moral reason to alert the child. This objection seems to stem from the worry that if an agent does not have a reason to perform a moral action, then he must not be morally responsible for his inaction.

Contrary to the claim of the objection being considered, internalism does not (in itself) deny the existence of some moral reasons/responsibility. The qualifier "in itself" is necessary because, as we will see, if internalism is combined with a theory of moral reasons and or responsibility, etc., then it may deny moral reasons and/or responsibility in cases where an agent is not capable of being motivated by a (perhaps otherwise[34]) moral reason^{E} to φ. But internalism^{H} does not, in itself, make any claims about moral reasons. The only way internalism^{H} has any implications for morality is if it is combined with an account of morality. For example, if a theory of morality claims that an agent does not have a moral obligation to φ if the agent does not *have* a reason to φ, then internalism would, when united with that moral theory, imply that the agent does not have a moral obligation if he is not capable of being motivated to φ. But, again, note that internalism *in itself* does not make a claim about morality.

To support my claim that internalism does not in itself deny moral reasons and/or responsibility, I will provide three considerations. The first is a reminder that, as I have argued,

[34] We may not want to say that it *is* a moral reason^{E} since the agent is incapable of acting upon it. *Perhaps* (and this depends upon our theory of morality) the putative moral reason^{E} is not a *moral* reason^{E} since the agent is incapable of acting upon it.

internalism^{H} takes the practical standpoint with respect to reasons for action. Therefore, it is concerned only with what action an agent should actually perform, given his limitations. And so it is only concerned with the reasons for action that an agent is actually capable of acting upon. If it really is the case that it is not within Andrew's capacity to be motivated to alert the child of the pothole, then he does not *have* a reason to alert the child. But that does not entail there is not a reason^{E} (which could be a moral reason^{E})[35] to alert the child. There may be a very good reason^{E} to alert the child, but it is not one which the agent has. But we are now likely to want an answer to the following question: What are the implications of internalism for agents who do not have a reason to act morally? Do they cease to have an obligation to perform the action? Do they fail to have any moral responsibility for not performing the moral action? As well, if we claim that an agent does not have a reason to act on a moral reason^{E}, does that preclude all negative moral judgment of the agent?

Those questions lead us to my second point: neither the formulation of internalism, nor the interrelationship principle which Williams claims is the fundamental motivation for it, make *any* claims about morality. In particular, they do not say anything about the moral status, obligations, responsibility, etc. of an agent who is motivationally incapable of acting upon a (perhaps otherwise) moral reason^{E}. Internalism itself leaves it a completely open question whether an agent who is incapable of acting upon a (perhaps otherwise) moral reason^{E} is morally responsible for not acting upon it. Whether the agent is morally responsible depends not upon internalism, but instead upon our account of what conditions are necessary for moral responsibility.

[35] So as to not take us too far afield, I will not make any claims about the nature of moral reasons to act, nor about their relation to non-moral reasons to act. Some moral theories may subsume all general reasons to act under moral reasons, e.g. utilitarianism. But, as I will emphasize in the second point, internalism does not make any claims about moral reasons, so we can leave it to accounts of morality to determine the nature of moral reasons to act.

To illustrate the silence of internalism on the issue of morality, etc., let me contrast it with an alternative understanding of internalism (one which no one accepts—at least as far as I am aware) which would have implications for morality—i.e., it would deny moral reasons, responsibility, etc. were an agent incapable of being motivated to act on a putative moral reason, obligation, etc. On this alternative interpretation, the interrelationship principle (which Williams said is the motivation for internalism) is identical with the ought-implies-can principle about morality.[36] It would be the general principle, and from it we would derive the internalist principle which pertains specifically to an agent's motivational capacities. The resulting internalist principle would be: an agent ought to φ for reason *r* only if he is capable of being motivated to φ for reason *r*. We can now consider the implications of the scenario with Andrew above. Since Andrew is not capable of being motivated to alert the child of the pothole, it is *not* the case that he ought to alert the child. Since the ought-implies-can principle identifies under what conditions an agent does not have a moral obligation to act—namely, when he cannot act—Andrew does not have a moral obligation to alert the child of the pothole. Here we see an obvious case where internalism *would* have implications, in itself, for moral reasons and obligations. If internalism were based on OIC, then it would imply that an agent is not morally responsible if he is incapable of being motivated to act on a moral reason^{E}. Since the agent could not act on the moral reason^{E}, it would not be the case that he ought to do so.[37]

But that interpretation of internalism is not the reasons^{H} interpretation. On the reasons^{H} interpretation, the ought-implies-can principle is *not* identical with the interrelationship principle.

[36] One philosopher, Jonny Anomaly, claims that the interrelationship principle is a principle about practical reasons which is analogous to, but not identical with, the ought-implies-can principle about morality. However, I did not mention his interpretation previously because, in essence, it also presupposes an instrumental or quasi-instrumental theory of practical reason. See "Internal Reasons and the Ought-Implies-Can Principle", *The Philosophical Forum* 39, no. 4 (2008): 469-483.

[37] But, it should be pointed out, compatibilists about free will and determinism are likely to reject this understanding of OIC.

The interrelationship principle—on the reasons^{H} interpretation—only claims that in order for an agent to have a reason to φ, it must be within his capacity to φ for that reason. It does not say *anything* about whether the agent has a moral obligation to φ or is morally responsible for not φ-ing. When internalism denies that an agent *has* a reason to perform some action, it is only denying that the action is one which reason requires the agent—as he actually is—to perform. It is not denying that the agent is morally responsible for his failure to be motivated by the reason. (Of course neither is it affirming that the agent is morally responsible.)

The third consideration which supports my claim that internalism^{H} does not *in itself* deny moral reasons and/or responsibility, is that there are accounts of moral responsibility which, when combined with internalism, entail that agents *are* morally responsible despite not having a reason to perform a moral action (and, alternatively, morally responsible for not *refraining from* an action which they *morally* should have refrained from). To illustrate this, we will consider two different accounts of moral responsibility. The latter account is perhaps more properly understood as an account of the appropriateness of the moral evaluation of agents. Both accounts allow that an agent can be morally responsible for not performing an action despite a motivational incapacity to perform it. In citing these accounts, I am not claiming that they are correct, nor that internalism supports either of them. Since internalism is silent on the issue of moral reasons and responsibility, it has nothing to say about their truth. Instead, my objective is only to demonstrate that internalism is compatible with accounts of moral responsibility which claim that agents are morally responsible even when they are incapable of being motivated to φ. That is, they are morally responsible even when they do not have a reason to φ

The beginnings of the first account of moral responsibility can be found in an article by Edward Sankowski. Although Sankowski was not addressing Williams's internalism, we can

draw on his insights to develop an account of moral responsibility compatible with internalism. Assume that "ought" really does imply "can" (understood in a libertarian sense). Suppose I get into a heated argument with a coworker, and in a burst of anger I *uncontrollably* blurt out a series of nasty and hurtful remarks. And by "uncontrollably" I mean that it was not within my capacity to refrain from blurting out the remarks. If "ought" does imply "can", then it seems I would not be responsible for those remarks because in that moment I could not have restrained my verbal attack.

However, according to Sankowski, I could be responsible for those remarks, if I am responsible for my emotions. In "Responsibility of Persons for Their Emotions", he argues that "if persons are responsible for a sufficiently wide range of behavior, and since their behavior can be oriented in intelligent ways to regulate certain of their emotions in the light of the rational assessment of emotions that they can make—it is quite defensible sometimes to apportion responsibility to a person for an emotion which he feels."[38] Therefore, even if at the time of the argument I could not control my emotions, I could still be responsible for my emotions if I am responsible for my emotional state. One possibility is that in the past, after a heated argument, I realized that I had an improper amount of anger and that I needed (at least) to try to lessen the amount of anger I feel in such situations. Suppose that after coming to those realizations I chose not to do anything about it. (I did not take any anger management classes for example). Then it seems I am responsible, at least to some degree, for the excessive anger I have in the present situation.

However, more must still be established if people are going to be responsible for their inability to act on a moral reason^{E}, since Sankowski focuses only on responsibility for emotions.

[38] Edward Sankowski, "Responsibility of Persons for Their Emotions", *Canadian Journal of Philosophy* 7, no. 4 (1977): 835.

He does not address whether we are also responsible for the actions or inactions which result from those emotions. But, we can see that it is only a short move from claiming that an agent is responsible for his emotions to the claim that he is responsible for the actions which those emotions produce or prevent. If, contrary to Williams's characterization of him, the cruel husband was able to recognize that he needed to treat his wife better, but knows that he cannot given his present emotional state, he could take steps to alter his emotional state. So, if he then deliberately chooses not to take those steps, knowing full well that he will most likely be cruel to his wife, then in the future if he is cruel to his wife, it seems that he could be responsible not just for his emotions but also his cruel actions or inactions.[39]

However, since emotions are only one component of an agent's motivational set, to make this account line up with internalism we would need to broaden our claim to include all elements of the agent's S. It would then become the following: *If* agents can be said to have the ability to control the contents of their motivational set, then they may have some responsibility for the elements of their S, and so perhaps for actions which they perform or are unable to perform as a result.[40] If this principle is true, then agents can be morally responsible for actions which they do not have reason to perform.

However, some will reject the idea that we can control the elements of our motivational set. Even if that is so, it does not *entail* that we are not responsible for our inability to act on a

[39] Knowing when an agent is responsible for actions which issue from his emotions is likely a very complicated matter. This is especially so since we often do not know what situations we will find ourselves in, and what type of emotional state we will need to be in to act morally. So, there may be a requirement that, if an agent is going to be responsible for actions or inactions which issue from his emotions, he must have been able to foresee that he would be in the situation, or at least was more than likely to be so. The case of the cruel husband seems to be a noncontroversial example of the latter. Given that he is married to his wife (and presumably lives in the same house) he can see that it is almost certain that he will be in situations where his emotions will impact his ability to act morally.

[40] Other considerations, one of which was mentioned in the previous footnote, might also be relevant to the assessment of an agent's moral responsibility for not having a reason to perform a moral action. We might need to determine 1) whether the agent could have foreseen that his influence on the shape of his S would result in his inability to act on that moral reason, or 2) whether the agent could have foreseen that the moral action would be one that he would have an opportunity to act on, and so on.

moral reason^{E}. Consider a second account of moral responsibility—or at least an account of the appropriateness of the moral evaluation or judgment of an agent—given by Simon Blackburn. In *Ruling Passions* he briefly addresses the debate over Williams's internalism. After noting the obscurity of the debate, he goes on to argue that even if an agent is incapable of being motivated to perform an action, moral appraisal of the agent can still be appropriate. In particular, the agent's desiderative profile may be otherwise than what it 'ought' to be. For Blackburn, the judgment that it is other than it ought to be does not entail that the agent-as-he-is is capable of changing his desiderative profile. Instead, it is only the judgment that the desiderative profile is a "bad" profile. The cruel husband who has no motivation to be nicer to his wife—and so does not have a reason to be nicer to her—can still be the subject of moral judgment for having a motivational set which results in treating his wife so poorly.[41]

Both of the above accounts of moral responsibility (or the appropriateness of the moral judgment of an agent) are compatible with internalism. They both claim that an agent can be morally responsible or merit moral evaluation for not performing a moral action, even if the agent did not have a reason to perform it. Again, I am not claiming that either of these accounts are entailed or even supported by internalism. Both of them could be false and internalism would still be true. Rather, I am only pointing out that internalism is compatible with them in order to demonstrate that internalism does not preclude an agent from being morally responsible for an action which it claims he does not have reason to perform.

III. A response to an objection to the formulation of the internalist thesis

[41] Simon Blackburn. *Ruling Passions* (New York: Oxford University Press, 2009), 264-6.

The last objection that we will consider is not an objection to the substance of the internalist position. Instead, it is an objection to the formulation of internalism as given by Williams. Also, it is not an objection held by anyone, as far as I am aware. But responding to it may help to dispel any lingering misunderstandings of internalism. According to Williams, the internalist formulation is "A has a reason to φ only if there is a sound deliberative route from A's subjective motivational set to A's φ-ing".[42] The problem with this formulation, the objector might claim, is that it does not actually rule out any reasons as ones that an agent has. In other words, although an agent is not capable of being motivated to φ for some reason *r*, and so does not have a reason *r* to φ, the formulation of internalism does not actually identify the reason as being one which the agent does not have. And that is because there is a sound deliberative route from any agent's S to any action which is most supported by reasons^{E}. So, even if an agent is not capable of being motivated to φ for reason^{E} *r*, the internalist requirement is met, because there is a sound deliberative route from the agent's S to the action. With this objection, note that the objector might agree with the "spirit" of internalism—that an agent who is incapable of being motivated to φ for a reason does not have a reason to φ—but he rejects the formulation of it, claiming that it does not actually rule out an agent having a reason^{E} as a result of lacking the motivation to perform the action.[43]

To show how this objection is supposed to work, I will utilize Michael Smith's conception of systematically justifying our desires. However, what follows is not Smith's objection. Instead, I am only using his conception of a process of systematic justification to

[42] Williams, "Postscript", 91

[43] However, it does seem to rule out an agent having a reason to φ when there is not a genuine reason^{E} to perform an action. Since there is not a reason^{E} to perform an action, there will not be a *sound* deliberative route.

illustrate the objection.[44] According to Smith, what there is reason^{E} for an agent to do is determined by what the agent would be motivated to do, were he fully rational. And, important to the objection being considered, given any agent's starting motivational set, there may be a process of deliberation which would determine what the agent would be motivated to do were he fully rational. Consider Smith's explanation of this process.

> Suppose we take a whole host of desires we have for specific and general things, desires which are not in fact derived from any desire we have for something more general. We can ask ourselves whether we wouldn't get a more systematically justifiable set of desires by adding to this whole host of specific and general desires another general desire, or a more general desire still, a desire that, in turn, justifies and explains the more specific desires that we have. And the answer might be that we would. If the new set of desires—the set we imagine ourselves having if we add a more general desire to the more specific desires we in fact have—exhibits more in the way of coherence and unity, then we may properly think that the new imaginary set of desires is rationally preferable to the old.[45]

So, from any agent's set of desires, it may be possible for the agent to deliberate about his desires to the more rationally preferable set. In fact, he may even be able to deliberate to the supremely rational set. If so, then there is a sound deliberative route from the agent's S to the action which the fully rational agent would be motivated to act upon. However, although there may be a sound deliberative route to the action which there is most reason^{E} to perform, it is not guaranteed that the agent would be motivated by that reason to perform the action. Given that the agent is less

[44] This objection could be illustrated by other conceptions of practical deliberation, but the type of deliberation considered is inconsequential to the objection, as well as to what is wrong with it.
[45] Smith, "Internal Reasons", 115.

than fully rational, it may be that his motivations (which are contrary to reason) preclude him from acquiring the motivations which a fully rational version would have. Smith gives as an example of this a squash player who has been defeated in a game of squash. Although a fully practically rational agent would be motivated to shake the hand of his opponent, the squash player, who is not fully practically rational, is so angry at losing that, were he to go to shake the hand of his opponent, he would become so enraged that he would end up slamming his racket into the face of his opponent. His less-than-fully-rational anger precludes him from actually shaking his opponent's hand.

This scenario appears to be problematic for the *formulation* of internalism. Since there is a sound deliberative route from the agent's S to the action which there is most reason^{E} to perform, it appears that the internalist requirement is met. So, the agent *may*[46] have a reason to shake the opponent's hand. And this is so despite the fact that the agent, because he is not fully rational and therefore does not have the same motivations as a fully rational agent would have, is not capable of being motivated to perform the action. One would think that this would be a prime example of a scenario where internalism would rule out the agent having a reason to perform an action. The spirit of internalism is that, in order for an agent to have a reason to φ, it must be within the agent's motivational capacity to φ for that reason. Therefore, Williams's formulation of internalism appears not to accurately capture the spirit of the internalist thesis. It does not rule out the agent as having a reason to φ, despite his inability to be motivated to φ for that reason.

Contrary to the claims of this objection, the problem is not with Williams's formulation of internalism, but with the objector's understanding of it. In particular, note that the internalist formulation requires that there be a sound deliberative route to *A's φ-ing*. It is not enough that

[46] We can only say that the agent *may* have a reason, since internalism is only a necessary condition for having a reason, not a sufficient one.

there is a sound deliberative route from the agent's S to the action in question. It is not as though the formulation merely says that there must be a sound deliberative route from *A*'s *S* to φ-ing. Instead, it requires that the route lead to *A*'s φ-ing. In other words, it is not enough that there merely be a sound deliberative route from the agent's S to the action. It must also be the case that the agent would, after following the sound deliberative route, be motivated to φ. If his present deficiency with respect to practical rationality entails that he would not have the motivation, then he does not satisfy the internalist requirement, and so does not have a reason to φ.

One further seeming problem for Williams's formulation may still seem to stand, however. Suppose that the agent in question is not fully practically rational, but the deficiency is related not to his motivations but instead to his theoretical rationality. In this case, he would have the motivation to perform the action which the sound deliberative route would recommend, but, due to his deficiency with respect to theoretical reasoning, he is not able to perform the deliberation which is necessary to recognize what there is most reason^{E} for him to do. In this scenario the agent does genuinely meet the internalist requirement. There is a sound deliberative route to *A's* φ-ing. That is, were A to deliberate soundly, he would be motivated to φ. However, since he does not have the rational capacity to φ, he cannot actually φ. It might seem then that the internalist formulation is flawed in that it does not preclude agents who are in this scenario from having a reason. Rachel Cohon raises an objection to this effect in her "Internalism about Reasons for Action".[47] (However, she takes it to be a problem for internalism itself, not merely the formulation of it.)

[47] Cohon, "Internalism", 273-8. Cohon claims that internalism is a requirement that there be *rational motivational access* to what there is reason^{E} for an agent to do. Since the agent in question is not capable of performing the necessary rational deliberation, he actually does not have rational motivational access to the reason^{E} even though he has the motivation.

This objection is correct that the internalist requirement is met. The internalist formulation does not deny that the agent has a reason. (But, since it only states a necessary and not a sufficient condition, it does not affirm that the agent does have a reason.) However, that is not a problem for Williams's internalism or his formulation of it. It is not a problem because internalism has a limited scope with respect to reasons^{H}, one which is restricted to the motivational capacities of agents. With respect to reasons^{H}, internalism is only the claim that an agent must have a motivation related to a reason^{E} in order for it to be one that the agent has. Internalism is not a requirement about the rational capacities of agents, nor their physical limitations. It does not claim that the agent must have the rational or physical capacity to act on a reason in order for it to be a reason^{H}. In the opening of his original account of internalism in IER, Williams noted that internalism is the claim that the lack of a motivation precludes an agent having a reason to φ.[48] He was not setting out to defend the more general principle that all incapacities to perform an action rule out an agent having a reason. However, he certainly thinks so. The interrelationship principle just is that more general principle.

That internalism has a limited scope also helps explain why Williams discarded a formulation of internalism which has not received much attention.[49] The formulation that we have been working with, "A has a reason to φ only if there is a sound deliberative route from A's subjective motivational set to A's φ-ing"[50], was originally stated in IROB and reasserted in "Postscript". However, in VRP, which came between those two articles, Williams provided a unique formulation of internalism. In VRP he claimed that the internal interpretation of the statement that "*A* has reason to φ" is that "*A* could arrive at a decision to [φ] by sound

[48] Williams, "Internal", 101.
[49] The lack of attention given to it I think is due to it not being readily apparent that it is substantially different from Williams's last formulation.
[50] Williams, "Postscript", 91

deliberation from his existing *S*".[51] This formulation requires not only that the agent have the motivation to φ, but also that he could arrive at the decision to φ. Given that this formulation also requires that *the agent could arrive* at the decision to φ, it would also rule out reasons which an agent could not act upon due to an incapacity related to theoretical rationality. Even if there is a sound deliberative route to *A*'s φ-ing, if the agent was not able to perform the deliberation, he would not have a reason to perform the action which would be supported by the deliberation. Since this formulation is more encompassing, we might think that it is a better formulation of internalism. However, given that Williams was concerned with the implications of the motivational incapacities of an agent on the reasons he or she has, it is understandable that he discarded this formulation. Internalism is only a claim about the effect of motivational incapacities to act on the reasons an agent has. It is not concerned with the effect of rational or physical (or any other) incapacities on an agent's reasons.

Summary

The truth of Williams's internalism has been the source of much debate. However, once we understand the nature of the internalist position—that an agent *has* a reason to φ only if the agent is capable of being motivated to φ for that reason—its truth seems obvious. Given that "has a reason" is defined such that its application is restricted to only those reasons which an agent is capable of acting upon, then the internalist claim that an agent who is incapable of being motivated by a reason^{E} to φ does not *have* a reason to φ follows as a matter of logic. A motivational incapacity is merely a particular type of incapacity.

[51] Williams, "Values", 109.

Despite internalism being true by definition, it has been important to respond to the objections in this chapter, largely because these objections rest upon a misunderstanding of internalism, one which my responses have aimed to dispel. For the most part, the objections rest upon an interpretation of internalism which takes internalism to be concerned with reasons^{E} instead of reasons^{H}. Since that that interpretation is mistaken, these objections are not relevant to internalism about reasons^{H}. Lastly, it should be noted that even if I have identified the correct interpretation of Williams's internalism and successfully defended its truth, the previous debates between internalists and externalists remain in place, insofar as those debates are predicated on a disagreement over whether the existence of *reasons*^{E} depend upon the motivational states of agents. Nothing that we have covered in our discussion of Williams's internalism directly impacts that issue, at least not in any noticeable manner.

Chapter 5: Williams's Internalism and the 'Morality System'

The preceding chapters have been concerned with providing a new interpretation of Bernard Williams's internalism, defending the accuracy of it against other interpretations, as well as defending the truth of internalism on that interpretation. That interpretation—as we saw—is the reasons^{H} interpretation. According to the reasons^{H} interpretation of Williams's internalism, internalism is the claim that a necessary condition for an agent *having* a reason to φ is that it must be within the agent's capacity to be motivated to φ for that reason. In this chapter I want to briefly recap both the nature of the reasons^{H} interpretation and the arguments for it being the correct interpretation of Williams's internalism**.** In the course of doing so I will also highlight what I think is the main cause of the confusion over the nature of Williams's internalism. Then, in the final section, I want to cover some new ground by drawing attention to a connection between Williams's internalism and his rejection of what he calls the 'morality system'. What we will see is that his views about each of them are driven in part by his strongly held beliefs that the motivational sets of human agents constrain what actions they are capable of performing, and so our theories of reasons for action and morality must take that into account.

I. Reasons^{H} and reasons^{E} interpretations of internalism

The reasons^{H} interpretation that I have been arguing for stands in contrast to the predominant interpretations of Williams's internalism. The predominant interpretations of internalism have taken it to be concerned with *reasons^{E}* for action. A reason^{E} for action is a consideration which counts in favor of performing an action. But, as we saw in Chapter 1, the predominant interpretations can be split into two different, more specific, interpretations. The

first, and most common, interpretation is that internalism is the claim that *reasons^{E} to act are constrained by the subjective motivational set of the agent.* As we saw, Shafer-Landau, Wallace, Hooker, and Finlay, amongst others, accept this interpretation (but most deny its truth). On this interpretation, if an agent does not have the motivation to φ, then there cannot be a reason^{E} for the agent to φ. The most common objection raised by this group against the truth of internalism (as they interpret it) is that it presupposes an instrumental or quasi-instrumental theory of practical reason. Only if we think that agents' motivations are the sole generators of reasons^{E} for action should we think that reasons^{E} for action are nullified by the absence of a motivation. As Shafer-Landau points out, even if someone has no motivation to develop a social life, it seems that there can be a reason^{E} for them to do so, given how much better the person's life would be.

The second group does not take internalism to rely on any theory of practical reason. And that is because they do not take internalism itself to claim that the lack of a motivation precludes the existence of a reason^{E} to act. Instead, on their interpretation, internalism is the claim that *a reason^{E} to act must be capable of motivating a fully rational agent*. In other words, if a consideration is to be a genuine reason^{E} for action, it must be one by which a fully rational agent would be motivated to act. If a fully rational agent would not be motivated by the consideration, then it is not a reason^{E} to act. As we saw in Chapter 1, those who accept this interpretation of internalism—namely, Korsgaard and Smith—accept that internalism is true. However, they reject Williams's claim that reasons for action are relative to an agent's subjective motivational set. In particular, they reject Williams's claim that the relativity thesis is essential to internalism. That is because—on their interpretation of internalism as stated above—it does not *necessarily* follow that what considerations a fully practically rational agent would be motivated by is constrained by the agent's present motivations. They would be constrained *only if* an

instrumental theory of practical reason is true. But, they claim, the latter is not a necessary part of internalism. So, internalism itself is not committed to the relativity of reasons^{E} to an agent's motivational set.

Despite the differences between the two groups, what is important for our present concerns is that they both agree that internalism is concerned with reasons^{E}. This of course contrasts with the reasons^{H} interpretation. The question now to be asked is why they came to the conclusion that internalism is concerned with reasons^{E}. Given that I am defending an interpretation which claims that internalism is concerned with reasons^{H} and not reasons^{E}, that question is especially poignant given that most, if not all, other interpretations have accepted the reasons^{E} interpretation. If the reasons^{H} interpretation is correct, how is it that so many readers misunderstood Williams?

II. The likely cause of the erroneous reasons^{E} interpretation

So how is it that most interpreters of Williams's internalism came to believe that it was concerned with reasons^{E}? I think it is fairly safe to say that the main source of confusion is Williams's claim that internalism is concerned with the rationality of the agent.[1] As I pointed out in Chapter 1, "rationality" is ambiguous between *theoretical* rationality, which is concerned merely with the beliefs of the agent, and *practical* rationality, which is concerned with the beliefs of the agent and perhaps also (depending on our theory of practical reason) the motivations of the agent.[2] Almost all readers of Williams have taken him to mean practical rationality. If we take the practical rationality reading, then it does seem that that internalism would be concerned with

[1] Williams, "Internal", 102-3.
[2] See pages 29-30.

reasons^{E}. Here's why. If we are concerned with an agent's practical rationality, it seems that we are concerned with the issue of what the agent would do were he fully practically rational.[3] And if we are concerned with what the agent would do were he fully practically rational, apparently we are *not* concerned with the reasons which an agent-as-he-is is *capable* of acting upon. That is, the consideration of whether an agent-as-he-is is capable of acting on a putative reason for action does not affect our judgment of whether there is a reason for performing the action. And that is because the action which an agent would do were he fully practically rational may be an action which the agent-as-he-is (who is less than fully practically rational) may not be capable of performing. But that must mean we are not concerned with reasons^{H}, since reasons^{H} are ones which an agent must be capable of acting upon. So, in short, if internalism is concerned with the practical rationality of an agent, then it must not be concerned with reasons^{H}. And so the most likely option is that it is concerned with reasons^{E}.

But the reasons^{H} interpretation takes Williams to mean theoretical rationality. It takes internalism to be concerned with what the agent would be motivated to do *were he to have correct beliefs*. The reason internalism (on this interpretation) is concerned with the agent's theoretical rationality is that it wants to determine the actual motivations of the agent. On the reasons^{H} interpretation of internalism, an agent has a reason to φ only if it is within the agent's capacity to be motivated to φ. But an agent who has a false belief (for example, a false belief concerning the proper means to an end) will *appear* to have different motivations than he actually has. Williams's gin and petrol example is a good example of this. The agent believes that there is gin in a glass and wants to mix it with tonic and drink it. But in reality what is in the glass is petrol. Since the agent is *not* motivated to drink *petrol* and tonic, but rather gin and tonic,

[3] If Williams did mean practical rationality, then many of the objections considered in Chapter 4 would be good objections to internalism.

his apparent motivation to drink what is in the glass does not count as a genuine motivation to perform the action. And since he does not have the motivation to drink what is in the glass, he does not meet the internalist requirement for having a reason to act. So the reason for being concerned with the agent's (theoretical) rationality is that internalism wants to accurately determine whether it is within an agent's capacity to be motivated by a putative reason to φ.

Unfortunately, nothing that Williams says about his use of "rationality" directly supports either the practical or theoretical readings. So the case for which reading is correct, and so whether internalism is concerned with reasons^{E} and reasons^{H}, must be built on other grounds. In particular, instead of arguing for which reading of "rationality" is correct and then concluding whether internalism is concerned with reasons^{E} or reasons^{H} on the basis of that conclusion, it is more productive to work the other way around. By considering Williams's manifold claims about internalism, we can assess whether those claims are more plausible on the reasons^{E} or reasons^{H} interpretation. And given which type of reason his writings (at least predominantly) support, we can then conclude that the corresponding reading of rationality is the correct one. If they support the reasons^{E} interpretation, then we should accept the practical rationality reading; if the reasons^{H} interpretation, then the theoretical rationality reading.

III. Why we should accept the reasons^{H} interpretation of internalism

I built the case for the accuracy of the reasons^{H} interpretation on the general grounds that it provides the most charitable interpretation of Williams. In Chapters 1-3 I provided several reasons for thinking that Williams's claims are most plausible on the reasons^{H} interpretation, and so therefore it is the most charitable interpretation. Here I only want to focus on the two most

compelling reasons (in my estimation) that were offered. The first reason was that only the reasons^{H} interpretation is consistent with two claims of Williams which I argued are most likely essential to the internalist position. The first claim is (R), that all reasons for action are relative to an agent's subjective motivational set. And the second claim is (N), that no particular conception of practical reason is presupposed by internalism. I supported the idea that the first claim is most likely essential by pointing out that Williams wrote that, with respect to internalism, *by definition* all reasons are relative.[4] And (N) was supported by drawing attention to Williams's repeated claim that internalism did not rely upon a theory of practical reason. (R) and (N) are therefore most likely essential to internalism. (I say "most likely" because there is a *chance* that Williams misspoke or would have revised his position if confronted with a problem with his view as he had presented it.)

As a first step in showing that *only* the reasons^{H} interpretation is consistent with (R) and (N), in Chapter 1 I showed that all of the predominant interpretations of internalism conflict with either (R) or (N). The first predominant interpretation, that *reasons^{E} to act are constrained by the subjective motivational set of the agent*, is inconsistent with (N). This interpretation, as most of its holders have pointed out, *does* presuppose a particular theory of practical reason. If an objectivist (value-based) theory of practical reason is correct, then that some action would promote value would generate at least a *pro tanto* reason^{E} to perform the action (even if not *most* reason^{E}). And this could be true even if an agent does not have a motivation related to the reason. The second predominant interpretation, that *reasons^{E} to act must be capable of motivating fully rational agents*, is inconsistent with (R). On this interpretation, internalism itself does not claim that all reasons^{E} to act are relative to an agent's motivations. Reasons^{E} are relative to an agent *only if* an instrumental or quasi-instrumental theory of practical reason is correct. But, on this

[4] Williams, "Internal", 102.

interpretation, neither the relativism claim nor those types of practical reason are essential to internalism. Hence, internalism allows for the possibility that (some or all) reasons^{E} are not relative to an agent's motivations. But that conflicts with (R). Therefore, we saw in Chapter 1 that there was some reason to think that neither of the two predominant interpretations is the correct interpretation of Williams's internalism. One additional reason for showing that both interpretations conflicted with a (likely) essential aspect of Williams's internalism is that it would open up the reader to the possibility that a new interpretation might actually be a better interpretation.

In the first section of Chapter 2 I provided a detailed explanation of such an interpretation—the reasons^{H} interpretation. According to the reasons^{H} interpretation, internalism is concerned with (not surprisingly) reasons^{H}, and not reasons^{E}. By definition, in order for an agent to *have* a reason to φ, it must be within the agent's capacity to φ. So, if an agent's limitations make it beyond his capacity to φ, then he cannot have a reason to φ. There may be most *reason*^{E} to φ, but the agent does not *have* a reason to φ. Internalism then is merely the assertion that motivational limitations have the same consequence as other limitations—they constrain the reasons that an agent *has*. Since the reasons^{H} interpretation of internalism is concerned with reasons^{H}, and *not* reasons^{E}, it is consistent with both (R) and (N). Internalism^{H} allows that any theory of practical reason can be correct. That is, it *allows* that any type of consideration can count in favor of performing an action and therefore generate a *reason*^{E}. Therefore it is consistent with (N). But since it claims that an agent *has* a reason to φ only if it is within his capacity to be motivated to φ, the reasons the agent *has* are relative to his motivational set. And so internalism^{H} is consistent with (R). So internalism^{H} is consistent with (R) and (N). And, since none of the other interpretations are consistent with (R) and (N), internalism^{H} is the

only interpretation which is consistent with both (R) and (N). That is the first of the two strongest reasons for thinking that the reasons^{H} interpretation is correct.

The second of the two strongest reasons is that the reasons^{H} interpretation provides the most charitable interpretation of Williams's argument against external reasons. This was demonstrated in the latter half of Chapter 2 and the first half of Chapter 3. In Chapter 2 I provided an explicit formulation of Williams's argument against external reasons, the accuracy of which I defended through textual support. The argument was based on Williams's claim that the fundamental motivation for internalism is the interrelationship between normative and explanatory reasons—what I have called the *interrelationship principle*. The argument was formulated at a level general enough to be compatible with most interpretations of internalism, as it included general terminology which could be specified in accordance with the various interpretations of internalism. In particular, the argument left the notions of "normative reason" and "possibility" ambiguous. Since the general formulation of the argument is valid, all that is left to consider is whether the premises and conclusion are true once the general terminology is specified in accordance with each interpretation. At the end of Chapter 2 I showed that when "normative reason" and "possibility" are specified in accordance with the reasons^{H} interpretation—"reason^{H}" and "within an agent's capacity", respectively—all of the premises and the conclusion of the argument are true, and so the argument is sound.

In Chapter 3 I finished the argument that the reasons^{H} interpretation provides the most charitable interpretation of Williams's argument against external reasons. In particular I showed that the predominant interpretations are each problematic in some way in which the reasons^{H} interpretation is not. To do so, I showed that each of them (as well as two other possible interpretations previously unconsidered) are problematic in that they 1) render the argument

unsound, or 2) render the argument sound, but only by presupposing a particular theory of practical reason, or 3) are inconsistent with the argument (by denying the truth of one the premises). Given these problems, the reasons^{H} interpretation is most charitable with respect to Williams's argument against external reasons—and so therefore the most plausible interpretation with respect to it.

So, the fact that the reasons^{H} interpretation of Williams's internalism is the only interpretation which is consistent with both (R) and (N)—which seem to be essential to internalism—and the fact that the reasons^{H} interpretation provides the most charitable rendering of Williams's argument against external reason, provides very good reason for thinking that the reasons^{H} interpretation is the correct interpretation.

And, since there are very strong reasons for thinking that internalism is concerned with reasons^{H}, we also have very good reason for thinking that, when Williams claimed that internalism is concerned with an agent's rationality, he meant *theoretical* rationality and not practical rationality. Internalism is not concerned with whether an agent would φ for *r* were he fully rational. Instead, it is concerned with whether an agent has the motivational capacity to φ for *r*, and to determine what the actual nature of the agent's motivations are, we need to adjust his apparent motivations for false beliefs (or perhaps the lack of true beliefs).

IV. Internalism and Williams's rejection of the morality system

Given that the correct interpretation of internalism is the reasons^{H} interpretation, there is every reason to think that internalism—on the reasons^{H} interpretation—is true. By understanding internalism to be concerned with reasons^{H}, internalism is seemingly true merely by the definition

of what it is to be a reason^{H}. A reason^{H} is such that an agent can *have* a reason to φ only if it is within the agent's capacity to φ. Therefore, if an agent has a limitation (e.g., a physical limitation) which prevents him from being able to φ, then he does not have a reason to φ. And internalism is merely the claim that psychological limitations—i.e. limitations on what an agent is capable of being *motivated* to do—have the same constraining effect on the reasons an agent has as other limitations. And in Chapter 4 we saw that the objections which have been raised against internalism do not apply to the reasons^{H} interpretation. They do not provide good reasons to think that internalism^{H} is false. So, we have very strong reasons to believe that the reasons^{H} interpretation is the correct interpretation of Williams, and that internalism is true.

So, why was Williams so concerned to defend its truth? Although any attempted answer to that question will certainly be speculative in nature, I think that it is fairly safe to conclude that Williams's concern was that philosophical theories were not making contact with real life. In particular, they were overlooking the fact that what actions *human* agents are capable of performing is affected by the contents of their motivational set. And so with respect to the internalism/externalism debate, Williams's problem with externalist theories of reason for action was that they failed to acknowledge the reality that motivational limitations of agents constrained what actions they had reason to perform. And it is this same concern which is the basis for Williams's rejection of what he calls the 'morality system'. In the rest of this chapter what I want to show is how Williams's *very* plausible belief about the motivational limitations of human agents leads him to reject the morality system.

1. Features of the morality system

Williams's notion of the morality system is most fully explained in the chapter, "Morality, the Peculiar Institution", in *Ethics and the Limits of Philosophy*. Given several of the tenets of the morality system (to be explained below), it is forced into the untenable conclusion that human agents never have psychological limitations which prevent them from performing an action which they have a moral obligation to perform. Importantly here, the morality system is not implying something about *ideal* moral agents, implying that *they* do not have psychological limitations. Rather, the morality system implies that *all human agents* are free from psychological limitations that prevent them from performing an action which they have a moral obligation to perform. It is because Williams believes that human agents *do* have psychological limitations that he objects to the morality system. Since the morality system implies that human agents do not have such psychological limitations, and because he strongly rejects that claim, he rejects the truth of the morality system.

To prevent any misunderstanding, it must be noted that Williams's objection to the morality system is not an objection to morality—when "morality" is understood generally. Rather, Williams's objection is to a particular conception of morality. However, Williams partially obfuscates this fact by making a distinction between ethics and morality, and then proceeding to raise various objections to morality as he defines it. So, in reading Williams, we must keep in mind that his use of 'moral' and 'morality' have a very narrow meaning. It should also be noted that Williams does not take the morality system to be identical with any particular philosopher's view (although it is perhaps most similar to Kant's view, it diverges from his view in some ways). Rather, he says that "[i]t is the outlook, or, incoherently, part of the outlook, of almost all of us."[5] It is for this reason that Williams claims much of the present discussion and debate within moral philosophy is predicated upon the morality system. That is, instead of

[5] Williams, *Ethics*, 174.

debating the merits of various ethical theories of which the morality system is just one theory, much of moral philosophy presupposes the truth of the morality system.[6]

As Williams sees it, morality (in the narrow sense) is a specific version of ethics. So how is it that morality is a specific type of the ethics? Williams claims that "Morality is distinguished [from the merely ethical] by the special notion of obligation it uses, and by the significance it gives to it. It is this special notion that I shall call 'moral obligation'."[7] According to Williams, the notion of *moral* obligation (in contrast to the more general notion of *ethical* obligation) has several features. Below I will highlight only those features most relevant to the issue of internalism. And, to show how those features imply that human agents do not have psychological limitations which preclude them from performing actions for which they have a moral obligation, at the conclusion of the explanation of each feature, I will provide a claim—one that is entailed by that feature—which will make that implication clear (once all of the claims are combined).

The first feature is that a moral obligation is a practical conclusion. It is a conclusion about what action an agent must perform or not perform. The importance of this feature is that it secures the connection between moral obligations and internalism, in that they are both concerned with actions. We can represent this feature with the claim that *if an agent has a moral obligation, then he has an obligation to perform an action.*

The next feature of moral obligation is that to have an obligation to perform an action it must be within the agent's power to perform the action. This feature is related to the fact that moral obligation is a practical conclusion. Williams writes that, according to the morality system, "[a]n obligation applies to someone with respect to an action — it is an obligation to do

[6] Ibid.
[7] Ibid., 174.

something — and the action must be within the agent's power."[8] The issue of what conditions must be met for an action to be within an agent's power is undoubtedly controversial. It is directly related to the ought-implies-can principle and so it carries with it all of the baggage from the dispute over the nature of that principle. But the reason that it must be within an agent's power to perform an action for which there is a moral obligation is due to morality's view of blame. According to Williams, "[t]here is a pressure within [the morality system] to…allocate blame and responsibility on the ultimately fair basis of the agent's own contribution, no more and no less."[9] It would (at least seem to) be unfair for an agent to be blamed for an action which was not within his power. Therefore, if an agent has a moral obligation, it must be within their power to perform it.

I think it is reasonable to conclude that the phrase "within an agent's power" is equivalent to the phrase "within an agent's capacity". So, to make the connection with internalism more clearly, we can restate this feature of the morality system as the claim that *if an agent has a moral obligation to perform an action, then it is within the agent's capacity to perform the action.*

The final feature of moral obligation is that it is inescapable. As Williams puts it, "[t]he moral law is more exigent than the law of an actual liberal republic, because it allows no emigration…"[10] We cannot opt out of our moral obligations or choose to no longer be a member of the moral world. Some (non-moral) obligations we can opt out of. For example, if I have an obligation to my employer to work fifty weeks a year, I can rather easily opt out of that obligation by resigning my employment. That is not the case with moral obligations. If there is a moral obligation to perform some action, then *nothing* can alter that fact—including the elements

[8] Ibid., 175.
[9] Ibid., 194.
[10] Ibid., 178.

in an agent's motivational set. That an agent's motivational profile cannot affect the existence of his moral obligations can be seen in Williams's statement that according to the morality system, "moral obligation applies to people even if they do not *want* it to", and an agent can be said to have a moral obligation "even if…they *want* to live outside [the moral] system altogether"[11] (emphases mine). This notion that an agent's lack of motivation to perform an action does not affect whether the agent has a moral obligation to perform it is crucial to the conflict between the morality system and internalism. So, we can say that one implication of this feature is that *an agent can have a moral obligation no matter what the elements of his motivational set are.*

2. The problem with the morality system

So how is it that these features together imply that human agents do not have psychological limitations which prevent them from acting on a moral obligation? To illustrate that they do, let us assume a scenario in which it is true both that an agent has a moral obligation to φ and that the agent is *not* psychologically capable of φ-ing (i.e. he is not capable of being motivated to φ). Drawing on one of Williams's examples, let us assume a scenario in which an agent has a moral obligation to keep a promise.[12] Let us say when Phil's daughter Mary started college, Phil promised that when she graduated he would hand over his vintage '66 Ford Mustang to her. However, in the years that it took for her to graduate he grew so attached to the Mustang that he cannot now be motivated to hand it over to her. (If you do not think this would count as a moral obligation, then you are free to consider another promise which would generate a moral obligation to keep it.)

[11] Ibid.
[12] Ibid., 192.

What we can see is that, taking the three features of the morality system as constants, we have to deny that Phil cannot be motivated to keep his promise—in order to avoid inconsistency. That is, we must claim that he *is* capable of being motivated to keep his promise. Under the assumption that Phil is *in*capable of being motivated to keep his promise, we might think that we should conclude that he does not have a moral obligation. But that would contradict the claim that moral obligations are inescapable—that an agent can have a moral obligation no matter what the elements of his motivational set are. Phil would have "escaped" his moral obligation as the result of his motivational set. So, we cannot conclude that Phil does not have a moral obligation to keep his promise. But if he has a moral obligation, under the assumption that Phil is incapable of being motivated to keep his promise, we now have a conflict with the second feature of moral obligations. Since Phil has a moral obligation to keep his promise, it must be within his capacity to be motivated to keep his promise. But that is just a rejection of our assumption that Phil cannot be motivated to keep his promise. Since the assumption that Phil is incapable of being motivated to keep his promise is inconsistent with the three features of the morality system, we must give up that assumption—if the morality system is to be consistent. And this conclusion generalizes to *all* moral obligations. We can never conclude that a human agent is incapable of being motivated to perform a moral obligation, because that would imply either that he has a moral obligation which he cannot meet, or, if we reject that implication and so deny that he has a moral obligation, that moral obligations are escapable.

We might be tempted to think that, because someone is not capable of being motivated to perform an action that we were therefore wrong to attribute a moral obligation to them in the first place. That is, we erred in ever attributing the moral obligation to them. However, even this would conflict with the claim that moral obligations are inescapable. It would imply that as long

as someone does not want to perform an action, then he does not have a moral obligation. Moral obligations *would* be escapable—by merely not wanting to perform them to begin with. So, in determining whether a human agent has a moral obligation to perform an action, the contents of his motivational set are irrelevant. But since according to the morality system the agent must be capable of performing his moral obligations, we have to conclude that human agents *are* capable of performing an action no matter the constitution of their motivational set. That is why Williams claims "there is a pressure within [the morality system] to require a voluntariness that will be total and will cut through character and psychological or social determination".[13]

But Williams rejects that idea. He writes, "[it] is an illusion to suppose that this demand can be met…This fact is known to almost everyone, and it is hard to see a long future for a system committed to denying it."[14] In other words, almost everyone knows that the composition of agents' motivational sets affect what actions they are capable of performing. And, given that recognition, we cannot (at least consistently) accept the truth of the morality system. One or more of its features must be given up. So, as I claimed, we can see that Williams's firm belief that human agents' motivations affect their capacity to act is the backbone of his objection to the morality system.

But there is also one further problem with the morality system, one which is connected to internalism. Since the morality system erroneously denies that the motivational sets of agents constrain what moral actions they are capable of performing, it implies that all human agents always *have* a reason to act morally, even when that is not the case. Williams writes:

[13] Ibid., 194.
[14] Ibid.

> When we say that someone ought to have acted [i.e. he had an *obligation* to act] in some required or desirable way in which he has not acted, we sometimes say that *there was a reason* for him to act in that way — he had promised, for instance, or what he actually did violated someone's rights. Although we can say this, it does not seem to be connected in any secure way with the idea that *he had a reason* to act in that way. Perhaps he had no reason at all. In breaking the obligation, he was not necessarily behaving irrationally or unreasonably, but badly. We cannot take for granted that he had a reason to behave well, as opposed to our having reasons for wishing that he would behave well…there are many different ways in which people can fail to be what we would ethically like them to be. At one extreme there is general deliberative incapacity. At another extreme is the sincere and capable follower of another creed. Yet again there are people with various weaknesses or vices, people who are malicious, selfish brutal, inconsiderate, self-indulgent, lazy, greedy. All these people can be part of our ethical world. No ethical world has been free of those with such vices… (italics are Williams's; underlining mine)[15]

It is worth noting here that we see Williams make a distinction between there being a reason and an agent having a reason.[16] In fact he appears to be using those phrases in a way (at least roughly) similar to my use of them. He seems to be indicating that an agent's promise to do something at least *counts in favor of* performing the action which would keep the promise. But, as he points out, there is not a *secure* connection between there being a reason to φ, and the agent *having* a reason to φ. Just because there is a reason to φ does not guarantee that the agent *has* a reason to φ, since the agent may not be capable of φ-ing.

[15] Ibid., 192.
[16] We saw that he also did this in his response to McDowell in "Replies".

And this is where a further problem results for the morality system. Since the morality system implies that human agents are always capable of performing moral actions no matter what elements are in their motivational sets, it implies that agents always *have* a reason to act morally.[17] But, given the very plausible claim that agents' motivational profiles *do* affect their capacities to act, the morality system is wrong on that point. It is not the case that human agents always *have* a reason to act morally.

Conclusion

What we have just seen is that for Williams a proper account of morality must take into account the actual nature of human beings. An account of morality which neglects the fact that human beings have motivational limitations which prevent them from performing some actions—even moral ones—does not provide an accurate account. That same line of thinking is embedded within Williams's account and defense of internalism about reasons for action. It is the basis for Williams's claim that an agent has a reason to φ only if there is a sound deliberative route from the agent's subjective motivational set to his φ-ing for that reason. Internalism is a claim about reasons for action which an agent *has*. Since, to *have* a reason to φ an agent must be capable of acting upon it, if an agent is not capable of being motivated to φ, then it appears obvious that he does not have a reason to φ. And that is all that internalism is claiming. It is merely the claim that a necessary condition for *having* a reason to φ is that it must be within the agent's capacity to be motivated to φ. For various reasons most philosophers have misunderstood Williams, thinking that he was claiming that *reasons*^{E} for action depend on an agent having a

[17] To be clear, it is not that the morality system denies the truth of internalism. That is, it does not claim that an agent can have a reason to φ even if not capable of being motivated to φ. Rather, what it says is that all human agents *are* capable of being motivated to φ, at least whenever φ-ing is something that is morally required.

related motivation—that is, they misunderstood him to be claiming that a consideration can count in favor of an action only if the agent is motivated to perform the action. Were that the case, the objection that internalism relies on a quasi-instrumental theory of practical reason would be correct, and the truth of internalism would be controversial. But, since internalism is only a claim about reasons^{H} and not reasons^{E}, internalism does not depend on such a theory. Nor should its truth be controversial—at least not once the nature of the view is properly understood.

Bibliography

Anomaly, Jonny. "Internal Reasons and the Ought-Implies-Can Principle" *The Philosophical Forum* 39, no. 4 (2008): 469-483.

Blackburn. *Ruling Passions.* New York: Oxford University Press, 2009.

Bond, E. J. *Reason and Value*. Cambridge: Cambridge University Press, 1983.

Cohon, Rachel. "Are External Reasons Impossible?" *Ethics*, 96, no. 3 (1986): 545-556.

------"Internalism about Reasons for Action" *Pacific Philosophical Quarterly 7*4, no. 4 (1993): 265-288.

Cullity, Garrett and Berys Gaut, ed. *Ethics and Practical Reason*. Oxford: Clarendon Press, 1997.

Davidson, Donald. "Actions, Reasons, and Causes" *The Journal of Philosophy* 60, no. 23 (1963): 685-700.

------*Problems of Rationality*. Oxford: Clarendon Press, 2004.

------"Rational Animals" *Dialectica* 36 (1982): 317-328.

Gert, Joshua. "Williams on Reasons and Rationality," in *Reading Bernard Williams*. Edited by Daniel Callcut, 73-93. New York: Routledge, 2009.

Finlay, Stephen. "The Obscurity of Internal Reasons*" Philosopher's Imprint* 9, no. 7 (2009): 1-22.

Foot, Phillipa. *Natural Goodness*. Oxford: Clarendon Press, 2001.

Goldman, Alan. "Reasons Internalism" *Philosophy and Phenomenological Research* 71, no. 3 (2005): 505-532.

Hooker, Brad. "Williams' Argument against External Reasons" *Analysis* 47, no. 1 (1987): 42-44.

Korsgaard, Christine. "Skepticism about Practical Reason" *The Journal of Philosophy* 83, no. 1 (1986): 5-25.

Kraut, Richard. *What is Good and Why*. Cambridge: Harvard University Press, 2007.

MacIntyre, Alasdair. *Dependent Rational Animals*. Chicago: Open Court, 2005.

McDowell, John. "Might There be External Reasons?". In *Mind, Value, & Reality*, 95-101. Cambridge: Harvard University Press, 1998.

Elijah Millgram, "Williams's Argument against External Reasons" *Noûs*, 30, no. 2 (1996): 197-220.

Nagel, Thomas. *The Possibility of Altruism.* Oxford: Clarendon Press, 1970.

Parfit, Derek. "Reasons and Motivation" *Proceedings of the Aristotelian Society, Supplementary Volumes* 71 (1997): 99-146.

Roberston, John. "Internalism about Moral Reasons" *Pacific Philosophical Quarterly* 67 (1986): 124-135.

Sankowski, Edward. "Responsibility of Persons for Their Emotions" *Canadian Journal of Philosophy* 7, no. 4 (1977): 829-840.

Scanlon. T.M. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1999.

Shafer-Landau, Russ. *Moral Realism*. Oxford: Clarendon, 2003.

Skorupski, John. "Internal Reasons and the Scope of Blame". In *Bernard Williams*, edited by Alan Thomas, 73-103. New York: Cambridge University Press, 2007.

Smith, Michael. "Internal Reasons" *Philosophy and Phenomenological Research* 55, no. 1 (1995): 109-131.

------*The Moral Problem*. Cambridge: Blackwell, 1995.

Sobel, David. "Explanation, Internalism, and Reasons for Action" *Social Philosophy & Policy* 18, no. 2 (2001): 218-235.

------ "Subjective Accounts of Reasons for Action" *Ethics* 111, no. 3 (2001): 461-492.

Wallace, Jay. "Three Conceptions of Rational Agency, *Ethical Theory and Moral Practice* 2, (1999): 217–242.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press, 1985.

------"Internal Reasons and the Obscurity of Blame". In *Making Sense of Humanity*, 35-45. Cambridge: Cambridge University Press, 1995.

------"Moral Incapacity" In *Making Sense of Humanity*, 46-55. Cambridge: Cambridge University Press, 1995.

------"Postscript: Some Further Notes on Internal and External Reasons". In *Varieties of Practical Reasoning,* edited by Elijah Millgram, 91-97. Cambridge: The MIT Press, 2001.

------"Replies". In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Edited by J.E.J. Altham and Ross Harrison, 185-224. Cambridge: Cambridge University Press, 1995.

------"Values, Reasons, and the Theory of Persuasion". In *Philosophy as a Humanistic Discipline*. Edited by A.W. Moore, 109-118. Princeton: Princeton University Press, 2006.