The Meaning of Integrity: A New Interpretation of Bernard Williams' Integrity Objection

The Meaning of Integrity: A New Interpretation of Bernard Williams' Integrity Objection

By Nicholas Francis Marshall, BA

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Art

M**cMaster University MASTER OF ARTS (2025) Hamilton, Ontario (Philosophy)**

TITLE: The Meaning of Integrity: A New Interpretation of Bernard Williams' Integrity Objection

AUTHOR: Nicholas Francis Marshall, BA (McMaster University)

SUPERVISOR: Dr. Stefan Sciaraffa

NUMBER OF PAGES: i–vii, 1–85

LAY ABSTRACT:

In this thesis, I argue that the famous "integrity objection" is not about defending an ethical person's right to pursue their deep attachments, relationships, and desires—as it is traditionally understood. Instead, it is about whether or not it is rational for an ethical person to be conflicted in some cases in which influential moral systems can easily generate a conclusion. For example, if a moral system endorses the conclusion that it would be best to execute one innocent person to save the lives of nineteen other innocent people, is it then irrational for an ethical person to have misgivings about having to carry out the deed themselves? Williams argues that this is not irrational, because we will have other ethical feelings and beliefs which cannot be reduced to systemic forms of morality.

ABSTRACT:

This thesis proposes a new interpretation of the famous "integrity objection", an objection to utilitarianism formulated in 1973 by the philosopher Bernard Williams. Other interpretations have claimed that the objection tries to show that utilitarianism unfairly alienates agents from their deepest desires, by demanding that they always pursue maximizing utility, and so neglect the projects and relationships which fill their lives with meaning. I argue that the objection is instead that utilitarianism—and other impartial moral theories—alienates agents from their own ethical judgements regarding those projects and relationships. The key difference is that I argue Williams is concerned only with the rational authority of "commitments", which are motives related to those personal ethical judgements, and not other private desires. Chapter One unpacks Williams' objection as arguing that moral theories ignore the rational salience of commitments, by demanding that an agent be apathetic to the reasons for action those commitments entail. Chapter Two critiques Samuel Scheffler's interpretation, who takes the objection to be that impartial morality is unduly demanding, because it restricts agents from fulfilling any of their desires or maintaining their personal relationships. I argue that this view cannot make sense of the way Williams illustrates the objection through his example cases, where agents are not subject to that restriction, and yet their integrity is nevertheless compromised. Chapter Three critiques Sophie Grace Chappell's interpretation, who argues that Williams claims that there can be no reason to follow the demands of impartial morality systems. I find that this is not Williams' claim, instead the objection argues that if there are reasons to follow the demands of impartial morality, those reasons do not therefore have the authority to make us ignore the reasons entailed by our commitments.

ACKNOWLEDGMENTS:

I am fortunate to have many excellent people to thank. I must first acknowledge the tremendous support I have received from my supervisor, Dr. Stefan Sciaraffa, professional and elsewise. Your clarity, attention, engagedness, expertise, and encouragement has led me through challenge after challenge, and only ever inspired me to improve.

I must also thank my family, especially my parents, whose endless enthusiasm and support has made it possible to pursue what I love. And I am indebted for the love and support of Anastasia, a debt I will be glad to repay for the rest of my life.

I am grateful for the dear friendship of many of my colleagues in the McMaster Philosophy department, including students, faculty, and staff. Most especially Colette Nguyen, Yasin Ali, Matthew Lam, Nigel Freno, Zachary Swirski, Firas Kayssi, Alexander Davidovic, and Logan Ginther.

And I must finally express my gratitude for Dr. Mark Johnstone, whose leadership and support has guided me in all my years at McMaster, for which I will always be grateful.

TABLE OF CONTENTS

Introduction

By *integrity*, Bernard Williams does not mean a virtue of being loyal to some value or creed. Integrity in this sense is not an admirable quality with a magnitude distributed variously among rational agents, such that some have more integrity than others, the way some agents may be more courageous than others. Instead, integrity is simply a part of agency, a property agents have that describes their relation to their own motives, including and especially that those motives endure. Agents, for Williams, have a set of motives which supply their reasons for action, and integrity is that one agent's motives are their own, and that the agent will be disposed to deliberate given those motives. The only people who would not have integrity are those whose motives are unstable to the point that we might begin to doubt that they are agents in the first place, perhaps as the result of unhealth. But so long as one deliberates given a set of motives which extends across time with some continuity, that agent has integrity.

In what follows, I defend this interpretation of integrity in various ways. Chapter One is an exegesis of the formulation of the integrity objection from "A Critique of Utilitarianism", where I define the objection as claiming that impartial moral systems characteristically ignore reasons derived from the motives of agents, to their detriment. While it may be desirable that a moral system encourages agents to set some motives aside, Williams takes exception with a type of motive he calls a "commitment". Commitments are the judgements an agent makes about what is right and wrong, and are understood as important considerations that the agent themself understands as involving the ethical. One could be committed to a moral system, to a principle, to a country, to a person, or to a practice, or some other object one values. For Williams, being

committed to something in this sense simply means that you will deliberate as though its object presents a pressing ethical concern, one you do not regard as trivial or unimportant[1].

Already, I have said that commitments can involve the ethical, and also that one can be committed to a moral system. As Williams eventually defines these terms, these are two very different claims. Williams cultivated a stable of technical distinctions among the sorts of motives and reasons which might play a role in an agent's deliberation. These distinctions are clarifying, and so, I will proceed in keeping with Williams' taxonomy.

Rational

An agent is rational when they follow reasons they actually have. These reasons are *explanatory*, in that their agent might actually become convinced of their validity, and so be motivated to act in the way picked out by that reason. An agent who cannot be convinced that they have a reason for action, and so could not act on that reason, does not have that reason. This is true regardless of the goodness of the action in question, or how prudent that action would be for the agent. This is why they are explanatory: they *could* explain an agent's conduct, if they deliberated in a way that led to them acting for that reason. But positing a reason that an agent could not be convinced of, and so can not follow, explains nothing of their conduct. Furthermore, these reasons are *normative*, in that if the agent becomes convinced that they have a reason, that belief must be sound for that reason to truly exist. The normativity follows from the relation between actions, and the motives which ground reasons for those actions. An agent's reason stands for a claim about an action the agent might undertake: that it satisfies, promotes, or protects some motive the agent has[2]. But an agent can be wrong about the relation between an action and their motives, if

[1] There will be more to say about what commitments are and how they influence deliberation in Chapter One, Section One.
[2] Bernard Williams, "Internal Reasons and the Obscurity of Blame" in *Making Sense of Humanity and Other Philosophical Papers, 1982 - 1993*, (Cambridge University Press, 1995), 39.

they (for instance) are under the false impression that some action will satisfy a desire they have. Hungry, I might think I have reason to walk to the store. But if the store is closed today, then my reason is not sound. If I know the store is closed, and walk anyways it would be irrational to undertake that action. So, an agent is rational insofar as they are responsive to the reasons they actually have. This is achieved through a) sound deliberation which identifies an agent's genuine reasons given their motives, and b) a willingness to act on those reasons.

Ethical

The ethical is a subset of the rational. In fact, a variety of non-ethical, or downright unethical considerations will be, in Williams' analysis, rational. Take the case of the neglectful husband, who has, after careful deliberation, no reason to treat his wife kindly. This agent, who is "ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal"[3], is nevertheless not irrational if it is true that they really have no motive which gives them a reason to treat their wife better.

The ethical is defined with reference to the content of an agent's motives. I take the distinction to follow this line: ethical motives arise from an agent's commitments, whereas non-ethical motives arise from desires for oneself. Williams contrasts ethical motives with those of the egoist: "The trouble with the egoist is not that it is desires that he expresses, nor that they are his desires - the trouble is that all his desires are for things for him."[4] Ethics, rather straightforwardly, is the business of evaluating what sorts of things agents are disposed to do, and what dispositions they have. He writes:

> "Some meta-ethical positions, including those that are, to my mind, the most plausible, have the following consequence: the characteristics that people acquire and exercise in ethical life and which are distinctive of it are not best understood on the model of cognitive or perceptual capacities, but rather on the model of dispositions of character. Let us call any view that has this consequence a *disposition view*."

[3] Ibid.
[4] Bernard Williams, "Egoism and Altruism" in *Problems of the Self : Philosophical Papers 1956-1972*, (Cambridge University Press, 1973), 260.

…

> "Granted a disposition view, it follows that dispositions have a certain kind of priority. This is not a priority within ethical thought: it is no more correct than it was before to say that all ethical value is the value of dispositions. What is true, rather, is that all ethical value rests in dispositions. Dispositions are basic because the replication of ethical life lies in the replication of dispositions. They are themselves among the objects of ethical evaluation, and are characteristics in virtue of which people themselves are thought to be better or worse; but, uniquely among those objects, they make the evaluation of all of them possible. In a certain sense, they give the value of those other objects, even though the value of those objects cannot be reduced to theirs."[5]

I take it that the evaluation goes something like this: the right dispositions are (at least) those that are not just in our own self interest, those are our commitments. So ethical motives are an agent's commitments. We will see below that commitments have some unique properties in deliberation, including a stringency such that they are inescapable—an agent cannot simply set their commitments aside in deliberation the way they can disregard a desire of theirs. For Williams, this is a key property of the ethical.

Moral

The moral is a subset of the ethical. Williams writes:

> "I am going to suggest that morality should be understood as a particular development of the ethical, one that has a special significance in modern Western culture. It peculiarly emphasizes certain ethical notions rather than others, developing in particular a special notion of obligation, and it has some peculiar presuppositions. In view of these features it is also, I believe, something we should treat with a special skepticism. From now on, therefore, I shall for the most part use 'ethical' as the broad term to stand for what this subject is certainly about, and 'moral' and 'morality' for the narrower system."[6]

In keeping with this distinction, I will use "moral" to refer to normative systems and systematic forms of ethics, noting consequentialist and deontological systems as paradigmatic examples. This essay understands the integrity objection as an objection to the primacy of the moral over the rest of the ethical. I will proceed using this moral/ethical

---

[5] Bernard Williams, "The Primacy of Dispositions" in *Philosophy as a Humanistic Discipline*, ed. A. W. Moore, (Princeton University Press, 2006), 74.
[6] Bernard Williams, *Ethics and the Limits of Philosophy*, (Harvard University Press, 1985), 6.

distinction, referring to Williams' target variously as "morality", "moral systems", "normative systems", "systematic morality", or "impartial morality", since I take it that the most problematic characteristic of moral systems is the way impartial deliberation excludes non-moral (but nevertheless ethical) considerations. Williams makes the moral/ethical distinction in the 1985 book *Ethics and the Limits of Philosophy*, but many of the earlier essays introduce what would be called ethical considerations under the name of moral considerations, namely "moral self-indulgence" and "moral incapacities". We will see that each of these concepts refer to considerations which have little to do with a normative system, and so should rightly be called "ethical self-indulgence" and "ethical incapacities". However, revising Williams' terminology would confuse the issue more than clarifying it. These make two notable exceptions to the rule.

In the formulation of the objection in "A Critique of Utilitarianism", Williams makes two substantial claims: first, that agents with commitments will not be able to deliberate impartially when they know that their commitments are relevant, and so some cases which will seem to be rather obvious when viewed impartially will not appear to be obvious for some agents actually involved. Importantly, this influence of commitments is not irrational, in that their commitments offer them legitimate reasons for action, and so an agent acting on those reasons is properly responsive to the reasons they actually have. And second, utilitarianism actually requires a conception of commitments to have a plausible understanding of what it means to promote happiness. Chapter One separates these claims, and takes the first claim to stand for what is properly called the integrity objection, following the tension between integrity and the command to deliberate impartially. The second claim augments the first claim's force against utilitarianism, by claiming that utilitarianism must assume a moral psychology which embraces commitments,

and yet fails to suitably accommodate them in deliberation. But the objection from integrity is a recurring theme in Williams' work, and exists in other formulations which can be used to better understand Williams' view. What is evident in all of these formulations is that Williams thinks his conception of commitments is independently plausible, and that the fact of integrity shows that a command to employ totally impartial deliberation is unjustified.

Williams characterizes the prospect of impartial deliberation as "absurd" in cases where an agent's commitments are at stake. Chapter Two works to show that Samuel Scheffler's influential interpretation of the integrity objection misinterprets that absurdity. For Scheffler, there is a burden in acting in ways which cannot be squared with one's deeply-held attachments, in abandoning things you find personally valuable in service of some "greater good". And this burden becomes absurd when it becomes exceptionally demanding—when it asks an agent to give up on too much of what they value. This leads Scheffler to develop a *hybrid theory* of morality, which combines consequentialism with an *agent-centered prerogative*, which is permission for agents to occasionally treat objects and relationships of personal value out of proportion with their value when viewed impartially. Scheffler's hybrid theorist is largely consequentialist, and would, for example, have an agent turn the trolley to save the five at the expense of the one, for purely consequentialist reasons. However, Scheffler's addition might grant the hybrid theorist permission to refrain from turning the trolley, if a loved one or dear friend would be put in harm's way.

The argument I forward shows that Scheffler misinterprets the absurdity in integrity cases, which is less about the effect on an agent who acts in ways which neglect their projects, and much more about the way an agent is expected to deliberate. What is offensive about impartial deliberation is not that it might reach the conclusion that one's projects are

insignificant, but that it proceeds from the assumption that they are only as good as they appear impartially. This is because, as one commitment among many, impartial moral systems do not have the authority to command an agent to disregard their other commitments, in part because of the inescapability of those commitments. And so, an agent's commitments will be in conflict with each other, putting them in an ethical dilemma. To command an agent to resolve an ethical dilemma by abandoning their commitment is to expect them to deliberate as though they were not themselves; this is what violates their integrity.

In the course of my argument, I make use of the claim from Chapter One that the integrity objection is one that Williams returns to, which can allow us to develop deeper insights into Williams' view. To this end, I help myself to the *rescuer* case. In the rescuer case, an agent's commitments give them a reason to act in a way which would save their spouse from a disaster, preferentially over strangers. And (Williams suggests), so might an impartial moral system. So, the same act is supported by reasons from the agent's commitment to their relationship, and the command of an impartial moral system. And yet, even in this case, the commands of morality threaten the integrity of the agent nonetheless. This is because an agent's integrity is not threatened by the prospect that they must act in a way which does not align with some of their pre-existing motives, it is threatened by the notion that the only motives available to them are those accessible from the impartial perspective. Even if morality is not *demanding* in commanding agents to act in ways which abandon their commitments, it nonetheless violates integrity by ignoring reasons from more personal sources: in this case, one's commitment to the welfare of their beloved. This is a feature of the objection that Scheffler's interpretation cannot explain.

Chapter Three takes up the relationship between the integrity objection and Williams' wider view of reasons and rationality. Sophie Grace Chappell understands the objection as presupposing two radically different sources of reasons: the motives which populate an agent's subjective motivational set, and morality. Chappell thinks that moral precepts are characteristically external to the desires an agent will typically have, and so we can conceive of moral precepts as demanding when they claim to supplant the motives we already have. She then argues that the integrity objection claims that impartial morality offers reasons which are unsound and uses the internal reasons thesis to explain why. In short, the internal reasons thesis claims that an agent will only have reasons which follow from their pre-existing motives, a set of desires which are crucial to the agent's ability to construct and live a life that is their own. This ability is, on Chappell's reading, their integrity. If the reasons offered by impartial morality systems are to be sound, they must at least be consistent with the motives which constitute the agent's integrity. But the moral reasons offered by impartial morality are not consistent with an agent's integrity, because the reasons offered by morality are prohibitively demanding, making it impracticable for an agent to live a moral life while also having a life of their own.

On Chappell's reading, the integrity objection explains why there are no reasons to follow the demands of impartial morality, relying on the internal reasons thesis to show that they are inconsistent with an agent's pre-existing motives. She frames the internal reasons thesis as an unstated premise in the integrity objection, which is meant to do the work of explaining what is so offensive about the demands of impartial morality. But, she worries that the internal reasons thesis cannot fill the mandate given to it by the integrity objection. I argue that Chappell's reading muddles the relationship between the internal reasons thesis and the integrity objection, and overstates the ambiguity in the internal reasons thesis. It can be tempting to read Williams as

though he is making a normative claim about impartial systems, as though their commands are inordinately demanding or unfair. The aim of this thesis is to defend a different interpretation of the integrity objection, which understands Williams as making a descriptive claim about agency. The accusation is not that impartial morality systems are guilty of *demandingness*, but that they are guilty of *distortion*, regarding the sorts of considerations available to rational agents. Chapter One defends this interpretation of the interpretation of the integrity objection. Chapters Two and Three contrast this reading with interpretations that take Williams to be making an essentially normative claim and not a descriptive one[7].

[7] Of course, Williams is loud and clear about the normative ramifications of this claim, but the argument turns on the plausibility of his conception of agency, and not an essentially normative appeal about fairness or demandingness.

Chapter One: The Meaning of Integrity

1. Introduction

There is a practice in moral philosophy to deploy counterexamples as evidence to the unsoundness of normative systems. The syllogism tends to follow like this:

> No plausible normative system will ever instruct an agent to ϕ, and in some circumstances, normative system X will instruct agents to ϕ, therefore normative system X is an implausible normative theory.

Here, use is made of some normative intuition, which features in the counterexample, and then seeks to show how in circumstances like the counterexample, the normative system in question cannot deliver the conclusion demanded by the moral intuition. What is presently important about this practice is that the integrity objection is not an objection of this sort. The conclusions we are supposed to draw from the objection is not that utilitarianism delivers unseemly verdicts. Rather Williams' integrity objection rests on a descriptive claim about agency, which casts doubt on the fittingness of the utilitarian style of deliberation. Specifically, Williams enlists what he calls "direct" utilitarianism as a foil in his initial statement of the integrity objection. As he characterizes this view, it holds that "the consequential value which is the concern of morality is attached directly to particular actions, rather than to rules or practices under which decisions are taken without further reference to consequence"[8]. Further, he takes satisfaction to be the maximand for this theory.

> It would trivialize the discussion of utilitarianism to tie it by definition to inadequate conceptions of happiness or pleasure, and we must be able to recognize as versions of utilitarianism those which, as most modern versions do, take as central some notion such as satisfaction, and connect that criterially with such matters as the activities which a man will freely choose to engage in.[9]

---

[8] Bernard Williams, "A Critique of Utilitarianism" from *Utilitarianism : For and Against* by J. J. C. Smart & Bernard Williams, (Cambridge University Press, 1973), 81.
[9] Williams, "A Critique of Utilitarianism", 85.

For the sake of brevity, I use the term "utilitarianism" to refer to this direct-satisfaction-consequentialism even though there are many other varieties of utilitarian theory.

The integrity objection is one line of argument from Bernard Williams' "A Critique of Utilitarianism", an essay which does not primarily attack the conclusions reached by utilitarianism, but rather pointedly asks the reader "do you really agree with utilitarianism's way of looking at the question?"[10]. Utilitarianism's way of looking at the question, according to Williams, involves a glib attitude to a subset of an agent's motivations he calls *commitments*.

Williams calls commitments "self-transcending social objectives which do not allow of trade offs."[11] As I understand this phrase, he means that commitments are motives which are other-regarding, and which appear to the agent as judgements about the ethical importance of some object, which are stringent in that an agent *will* deliberate as though they are pressing ethical concerns, even if it is in conflict with another commitment. For Williams, any agent disposed to some form of ethical thought will have commitments, which will influence the way they deliberate their way through ethical dilemmas and conflict. But it is not straightforward to see what that influence will look like. Compare commitments with desires. A commitment will differ from a desire in that the agent will not be able to take seriously the prospect that the commitment does not represent a legitimate reason for action, based on the importance of its object (from that agent's perspective). This is why there can be no trade offs: the object appears to the agent as ethically significant in its own right, and is therefore not a fungible source of some other value.

[10] Ibid, 78.
[11] Ibid, 146.

Some desires are fungible this way. I want a nice meal, and open the menu to two favourites: pork schnitzel and spaghetti carbonara. Both serve the higher-order desire of enjoying a nice meal, but in ways which are replaceable by another. So, my deliberation is only instrumental: carbonara is a little too rich tonight, but the schnitzel is just right. One is traded off for another, and without the loss of some value, as the value of either option is wrapped up in achieving the end of having a nice meal. But what if I am a vegetarian? What's more, my vegetarianism is ideological; it is a project which follows from a commitment involving the value of animal welfare. Once more I glance at the menu, and see that both contain pork. My deliberation can go only one of two ways. Another agent might think something along the lines of "Sod it, I have been a vegetarian all day already! Hogs be damned, I'll be having the schnitzel". But that could not be me. I have a commitment, and as Williams understands commitments, that means it cannot be ignored in this way, and so the agent who has this thought does not have the commitment to animal welfare.

But this does not mean that commitments will always win out in deliberation. In the cases Williams makes use of, agent's commitments will often be in conflict—this is what will be dilemmic about ethical dilemmas, that two exclusive commitments make it so that the agent cannot act losslessly, as any course of action will neglect at least one of their values[12]. Nor can it be said that at least one commitment will always win out over an agent's other desires. For Williams, it is perfectly intelligible that an agent can arrive at a conclusion regarding what they ought to do, ethically, and then to ask a further question of what they ought to do, practically. Indeed, this prospect is preferable to views which proclaim the "necessary supremacy" and

[12] I will therefore refer to what Williams sometimes calls "moral dilemmas" with the term "ethical dilemmas", as dilemmas emerge from conflict between commitments, which need not reference a moral system.

"ubiquity" of the ethical[13]. So Williams cannot say that an agent will always act on a commitment if it is relevant in deliberation, because that, too, would proclaim the supremacy of the ethical.

So if a rational agent will not always choose to act on reasons following from their commitments, what is their special place in deliberation? It is that those reasons will always be seen as legitimate candidates for action; they cannot be excluded as ineligible grounds for action by some other consideration. Some values can be excluded in this way. In our example at dinner, the vegetarian might be able to stomach sidelining their enjoyment of a nice meal, and consider the tastiness of the schnitzel to be an ineligible ground for their choice given their feelings about its origin. The point is clearer in more serious cases. A burning building sputters with reasons to avoid it, both for reasons connected with desires not to be in pain and motives connected to prudence. But for some, knowledge that their child resides helplessly somewhere on the second floor causes the motivational force of these reasons and motives to fade away. When that agent chooses, it is not as though their own welfare doesn't matter, but it just does not have a seat at the table when they choose what they are going to do. Naturally, if the circumstances were reversed, and the *danger* cooled their concern for their *child* into apathy, they cannot be said to be committed to their child in Williams' sense.

What's more—and this really is the central claim of the text—the influence of these commitments over an agent's deliberation is *legitimate*, in that it is not regrettable. An agent might be admonished for choosing selfishly, or shirking their duty out of cowardice. However,

[13] Bernard Williams, "Ethical Consistency" in *Problems of the Self : Philosophical Papers 1956-1972*, (Cambridge University Press, 1973), 185. In this paper, Williams objects to the ubiquity of the "moral", but keeping in mind the later moral/ethical distinction, I have changed the verbiage to "ethical", since the disposition he worries about is not necessarily tied to a normative system.

Williams thinks that agents develop commitments naturally and unavoidably. And what's more, necessarily, if they are to secure a meaningful, satisfying life.

Utilitarianism, however, fails to recognise the importance of commitments and is subsequently forced to either label those commitments irrational and regrettable shortcomings which ought to be discouraged and ignored in the utilitarian calculus, or to treat them as fungible opportunities to promote the agent's satisfaction. But Williams argues that utilitarianism must recognize that commitments are not irrational, lest this theory be forced to label every motive (save the motive to impartially promote satisfaction) as irrational. Since agent's commitments are contrasted, in Williams view, with an agent's egoist motives, this would mean that only egoistic motives are fed into the utilitarian calculus (along with the commitment to utilitarianism itself), and so utilitarianism would only be capable of registering changes in satisfaction that result from egoist motives. But this is an impoverished view of human satisfaction, since an agent's commitments are amongst the most dear and important pieces of a rich and valuable life. However, if utilitarians concede this point and admit that agents should promote motives which it views as irrational, they might only defend the principle of utility by acknowledging that commitments are the sort of thing that ought to afford a significant amount of weight in the utilitarian calculus. This would be admitting that commitments are the indispensable fixtures in an agent's life satisfaction, key determinants in their quality of life. But commitments, Williams claims, are more than important opportunities to promote the agent's utility, they are characteristically constitutive of an agent's ethical point of view, and if they are present in an agent, simply cannot be crowded out by the overwhelming normative force of the principle of utility. The descriptive claim Williams hopes to stress is that their significance is reflected not only in the worth of a commitment's satisfaction, but in its effect on an agent's motivation *from*

their own perspective. This is why, in the objection's culmination, it is *absurd* to demand that an agent sets them aside: they represent a competing point of view that the agent has taken up insofar as they have that commitment. And insofar as they do have that commitment, they are not purely a utilitarian, and so are *unable* to give in to its demands to view their own projects only instrumentally to the demands of utility maximization.

2. Negative Responsibility

Williams illustrates the tension between utilitarianism and integrity by raising two cases involving negative responsibility. Indeed, Williams says that utilitarianism is committed to a "strong doctrine of negative responsibility"[14] due to its singular occupation with states of affairs. Since the moral theory only registers differences between the quality of actions in terms of their consequences on the states of affairs, it does not admit of a difference between doing or allowing harm (or benefit), and so—if all else is equal—an agent is equally responsible for the harm that they inflict, and the equal harm another person inflicts that the first agent could have prevented[15].

In the two cases offered by Williams, circumstances out of the agent's control force them to choose between acting on their various commitments, and what utilitarianism calls for given the doctrine of negative responsibility. According to Williams, utilitarians will find determining what to do in these cases rather obvious, and this is seriously disconcerting. The doctrine of negative responsibility explains both of these claims. For, it will be seen that any difficulty in reaching a decision in these cases which come from an agent's commitments cannot satisfactorily be described as differences in utilitarian consequences, and so cannot be registered by utilitarianism as legitimate difficulties. The tension is between the serious (and intuitively

[14] Williams, "A Critique of Utilitarianism", 95.
[15] Ibid.

appealing) reservations felt by the protagonists of the cases, and the apathy of the impartial normative system which would ignore those reservations.

George

George is a chemist without a job, and whose family is in desperate financial need. After a period of serious difficulty in finding work, an older colleague reaches out to George with an offer from a laboratory which researches chemical and biological weapons. George is opposed to chemical and biological warfare, and so moves to decline the offer, before the colleague points out that George's refusal will not prevent the research from being continued. Further, he hopes George will accept the offer in part to prevent another candidate from getting the job, a candidate without George's scruples, who will likely pursue the research more effectively, producing more dangerous weapons.

Jim

Jim is a traveling researcher, who by happenstance encounters a town controlled by a rebel warlord. The warlord was set to execute twenty randomly selected civilians for political reasons, before Jim appeared. The warlord, seeking the legitimacy offered by collaboration with a foreign national, offers Jim the following deal: if Jim executes one of the civilians himself, the other nineteen will go free. Jim, knowing he cannot overpower the warlord and his soldiers, must either accept the offer, or refuse, allowing the warlord to carry on with the execution of all twenty civilians. The prisoners' families beg George to take the deal.

In each of these cases, Williams claims that utilitarians would say that the answer is obvious, since one option would clearly produce the greatest overall satisfaction, regardless of the private reservations of the agent[16]. It is easy to see that this is true. Nineteen lives would be

[16] Ibid, 99.

saved by Jim's intervention, set against the cost of his own involvement in the execution (presuming he is motivated to resist the killing of an innocent prisoner). If Jim refuses, nobody seems to benefit other than Jim. As Williams claims, a utilitarian would likely find this case quite easy. So too for George. George's family only stands to gain from his involvement with chemical biological research. Moreover, if his involvement serves to make the research itself less effective compared to the involvement of the other candidate, George might see that the only thing counting against his involvement is his own misgivings. A utilitarian might count this feeling as a reduction in satisfaction which counts against the choice of taking the job, but how could this feeling outweigh the benefit gained by his family, or the lives which might have been saved by his slowing of the research? As Williams points out: "they are after all (and at best) one man's feelings"[17]. And so it seems that without some special further stipulation, Willaims is correct in claiming that on the utilitarian view the answer to these cases is rather obvious.

Williams thinks we should be suspicious of the claim that these cases should be obvious for Jim and George. Not only is he hesitant to say that any answer in these cases could be *right*—let alone *obviously so*—he suspects that utilitarian analysis goes awry when "it cuts out a kind of consideration which for some others makes a difference to what they feel about such cases: a consideration involving the idea… that each of us is specially responsible for what *he* does, rather than for what other people do"[18] and fails to "coherently describe the relations between a man's projects and his actions"[19]. This shortcoming is caused by a procedural neglect of an important consideration on the part of utilitarianism, which cannot be registered when an agent's deliberation is tied only to the maximization of satisfaction.

[17] Ibid, 102.
[18] Ibid, 99.
[19] Ibid,100.

At this point in the argument, one might expect Williams' crack at describing "the relations between a man's projects and his actions". But Williams will not directly provide such a description. He will, however, defend some essential features of that relation from utilitarian dismissals of its importance. Indeed, utilitarians cannot account for this relation beyond the consequences that an agent's conduct brings about, and so utilitarianism must dismiss misgivings rooted in some other features of that relation as fatuous, or corruptive. If Jim complains that his commitment—expressed via misgivings—represents some critical feature of agency, utilitarian philosophers might attempt to dismiss such misgivings as Jim's *moral self-indulgence*[20].

3. Moral Self-Indulgence

As Williams portrays the possible objections utilitarians might make, Jim's misgivings are self-indulgent because they express a hesitancy to *do what must be done* if doing so comes at the expense of their own character. For instance, we might understand George as having a commitment which makes him see participation in research into chemical and biological weapons as ethically wrong, *regardless of consequences*. For the utilitarian, the ethical worth of such research is determined by the consequences that would follow from the research, and so any feeling that such research is *inherently wrong* is simply mistaken.

The self-indulgence objection must amount to the claim that Jim or George's deliberation is polluted with an undue concern for how their sense of their own conduct will fare given their decision, undue because of the weight it is afforded weight over the livelihoods of those victimized by their choice[21]. This allows the utilitarian to declare that any seeming obscurity in the case is the result in the agent's confusion about what moral deliberation aims at, and conversely recast the so-called *obviousness* of the case as the result of a high-minded clarity.

[20] Ibid, 102.
[21] Ibid.

According to Williams this response is the result of an important distinction utilitarians fail to make. It is plausible that in some cases, an agent will elect to do what they know is not right out of some undue self-indulgent concern, whether it be a draft-dodging sluggishness or conceited attention to their own reputation. These motives, which are reasonably viewed with suspicion, might even follow from a project an agent has or holds dear. However, one class of projects, which Williams terms "commitments", are distinguished from other motives because they involve an agent's ethical motivations, they are emotional reflections of their view of right and wrong[22]. It would be incomplete to say that the difficulty for Jim is that the utilitarian thing to do involves some unpleasantness, as that unpleasantness derives from Jim's sense that the killing of an innocent and helpless prisoner is itself wrong. The same is true for George, it is not just that he does not want to participate in the research program, it is that he feels there is something unethical about taking part in the program.

Of course, the utilitarian cannot agree with Jim or George: they must deny that there is anything wrong with killing the prisoner or taking part in the research, because utilitarianism can only explain the rightness or wrongness of actions in terms of their consequences, and they are assured that doing so will lead to the best overall outcomes. The charge of self indulgence, then, must be understood as an invitation "to consider [the] question from the utilitarian point of view"[23]. If Jim or George harbor some not-strictly-utilitarian disposition regarding the ethics of their choices, they therefore are not thoroughgoing utilitarians, then so their misgivings are not evidence of self-indulgence. They simply betray that they are not throughgoing utilitarians, the feelings are "indications of what [they] think is right or wrong"[24]. And the point here is not only about the content of Jim or George's character; Williams hopes that *we* will sympathize with his

[22] Ibid, 103.
[23] Ibid, 102.
[24] Ibid, 103.

protagonists, and find the utilitarian charge of self-indulgence "unsettling", because "we are partially at least not utilitarians, and cannot regard our moral feelings merely as objections of utilitarian value"[25]. If we have ethical intuitions which do not align with the injunction to maximize satisfaction, then the self-indulgence charge cannot cow us into abandoning those feelings for utilitarian orthodoxy. It is an appeal *by utilitarians for utilitarians*, one that could only work to restrain likeminded comrades who have fallen out of line.

4. The Importance of Commitments to Satisfaction

How might a utilitarian regard these countervailing commitments? For the convinced utilitarian, commitments and their associated feelings might be considered errors introduced by the perspective of the particular agent who holds to them. But to think of commitments as implicit claims about the objective value of objects misses the point.

Janine lays flowers at David's grave. One might ask her why she does that, to which she responds "to honor him". How does laying flowers honor a dead man? Why honor him anyways (he's not likely to notice)? These questions, and the lines of questioning which might follow them, are not likely to terminate in some claim about the objective value of the practice. Or at least, they do not need to for the practice to be intelligible. Instead, Janine honors David because she developed a commitment to him when he was alive, and Janine lives in a cultural context in which that is how a commitment to a loved one motivates one to act. And while this might offer a sociological explanation for her act, it will not explain why it is rational, in Williams' view, because it does not explain how the action relates to her contingent motives. It is her motives that make it rational, and so the act is rational if she has the requisite motive. But what of the motive; is it rational to be motivated to honor the dead in this way? For Williams, this is not the right sort of question. It is not rational or irrational to have one motive or another, because rationality is

[25] Ibid.

between an agent's motives and their conduct, and not between an agent and their motives. An agent is not irrational for having the wrong motives, but they are for following reasons which are wrong for the motives they have.

Nevertheless, commitments do make a claim on the agent who has them, in deliberation. There is a motivational pressure on Janine to honor David, and so in Williams' view she has a reason to do so. Utilitarians might doubt this claim. And indeed, Williams notes how this stipulation would handily avoid embarrassing conclusions utilitarians might otherwise be forced to reach. For instance, in a racially polarized society, where the hatred of the minority population makes it so that their mere existence causes a great deal of suffering for the majority, it stands to reason that utilitarians could be forced to endorse the "removal" of that minority, as their misery would be outweighed by the relief felt by the majority[26].

This example is not introduced to argue that utilitarians naturally endorse ethnic cleansing. The point here is that the pain of the majority which function as inputs into the utilitarian calculus, insofar as they include the bigotry of the majority, are the result of irrationality in the majority. The utilitarian might be motivated to explain how they are not committed to endorsing the position Williams argues them into, but they can only do so by appealing to the quality of the resultant consequences. And so, it would be convenient if they could exclude the consequences which seem the most ghastly, by way of their being irrational. And of course, the racist feelings of the majority *would be* irrational, insofar as they express a commitment about the suitable treatment of the minority, a commitment which is inconsistent with the principle of utility, and the fact that their misery is that of another person[27].

---

[26] Ibid, 105.

[27] And even another creature. This point is worth making to show that this argument is resistant to circumstances where the majority harbours wildly dehumanizing views about the minority; even if members of the majority are under the impression that the minority are in some way subhuman, the utilitarian majoritarian must still regard their suffering as an evil.

Such a result would protect the utilitarian from the claim that they must endorse the ethnic cleansing from Williams' example above, and would at the same time settle the question of the *obviousness* of Jim and George's circumstances. Jim and George's reservations would be irrational from the utilitarian point of view, and so irrelevant to the utilitarian calculus.

Williams does not think the utilitarian is entitled to this convenient result. This is because, as he thinks, that utilitarianism's pursuit of satisfaction is meaningless[28] unless the philosophy acknowledges and endorses an agent's engagement with projects they personally regard as intrinsically valuable. If utilitarians hold that the only pleasures worth having are those which are rationally consistent with the principle of utility, then they are committed to a psychologically shallow view of happiness and its attainment. But they need not be: utilitarianism could embrace commitments and other sorts of personal projects to become a more plausible moral theory, but at the cost of acknowledging that Jim and George need not deliberate as though they were thoroughgoing utilitarians; by ceding authority over an agent's deliberation—in part—to the projects an agent involves themselves in.

Williams goes about showing how necessary agents' commitments are to the utilitarian ambition by getting clear on what the utilitarian ambition actually is. He writes:

> What projects does a utilitarian agent have? As a utilitarian, he has the general project of bringing about maximally desirable outcomes; how he is to do this at any given moment is a question of what causal levers, so to speak, are at that moment within reach. The desirable outcomes, however, do not just consist of agents carrying out that project; there must be other more basic or lower-order projects which he and other agents have, and the desirable outcomes are going to consist, in part, of the maximally harmonious realization of those projects…[29]

The most satisfaction can only be achieved through engagement with the values agents actually have. At this point in the argument, Williams is concerned with an agent's *projects*, which are

[28] In more sense than one. Williams hopes to show that aiming towards happiness without pursuing other projects which are regarded as intrinsically valuable undermines the philosophy's ability to meaningfully guide action, which is attributed to the fact that those projects are necessary for an agent to experience their life as *meaningful*.
[29] Ibid, 110.

ongoing endeavors undertaken by agents by a series of actions across time. If writing a sentence in an important essay is an action, then it might help constitute the project of passing Philosophy 101, or the project of achieving a major in philosophy, or even the more lofty ambition of being a success. Often, an agent's projects ground their motivation to undertake particular actions, insofar as they are understood as contributing to the project, and so move the agent to do something which they might otherwise have had no desire to do.

One type of project is motivated by an agent's commitments. As we have seen, commitments involve an agent's beliefs about right and wrong, but projects might not involve this motive. An agent's projects can be self-interested, or malevolent, or prudent, or partial. The point here is that utilitarians must endorse projects which are not already ethical, if their philosophy is going to bear on actual practices. Williams provides a list of the sorts of projects that utilitarianism must promote as instrumental goods: "the obvious kinds of desires for things for oneself, one's family, one's friends, including basic necessities of life, and in more relaxed circumstances, objects of taste"[30], as well as "projects connected with [an agent's] support of some [political] cause… Or there may be projects which flow from some general disposition towards human conduct and character, such as a hatred of injustice, or of cruelty, or of killing"[31].

This all might seem well and good to the utilitarian, as it is easy to see that if a utilitarian hopes to promote the satisfaction of the other, it is necessary that the patient is psychologically receptive to some act which will make them satisfied. They might object, however, to some projects, on the grounds that they are derived from commitments, and therefore in conflict with the utilitarian point of view. Williams represents the worry rather obscurely:

> For this project will be discounted, presumably, on the ground that it involves, in the specification of its object, the mention of other people's happiness or interests: thus it is

[30] Ibid.
[31] Ibid, 111.

> the kind of project which (unlike the pursuit of food for myself) presupposes a reference to other people's projects.[32]

The thought is that commitments involve a view of how an agent is supposed to regard the interests of others. For instance, an agent's commitment to the welfare of their mother might influence them to take a high paying job, and pass up an opportunity to do good for anonymous strangers at a low-paying charity. But the utilitarianism Williams takes as his target purports to offer a complete view of how one ought to treat the projects of others, and provide an exhaustive list of an agent's reasons. And so, insofar as these commitments are not identical with the utilitarian project, they are irrational, as it would deny that the agent has any special reason to treat their mother with impartial favor. So, the utilitarian might again reaffirm the convenient stipulation that projects which come from commitments are not the sort of projects which the utilitarian calculus should work to satisfy.

But this stipulation may not be so painless as utilitarians hope, says Williams: "But that criterion would eliminate any desire at all which was not blankly and in the most straightforward sense egoistic"[33]. Thus, the throughgoing utilitarian could only have motives which are elementally egoistic, or messianically impartial. He continues:

> But the utilitarianism which has to leave this hole [between egoism and impartiality] is the most primitive form, which offers a quite rudimentary account of desire…Utilitarianism would do well then to acknowledge the evident fact that among the things that make people happy is not only making other people happy, but being taken up or involved in any of a vast range of projects[34].

If the ambition of utilitarianism is to produce the greatest amount of satisfaction, then it must permit agents to engage with their commitments, as these are among a person's most cherished projects. If utilitarianism is to be plausible, especially the satisfaction-consequentialism which

---

[32] Ibid.
[33] Ibid.
[34] Ibid, 112.

Williams takes as his target, then the theory must say that the satisfaction of agent's commitments should be promoted.

At this step, the utilitarian might again concede and agree with Williams that the normative system must reckon with the importance of each agent's commitments. This means that agents' commitments would be amongst the desirable consequences endorsed by the normative theory. And yet, it represents this result in deliberation by including them as input into the utilitarian calculus, just as any other desirable consequence would be. This seems a natural step, as it retains the utilitarian deliberative posture whilst acknowledging the importance of commitments in the happiness of agents. But Williams has identified two problems which would still apply to this considered view. First, Williams thinks it is at least conceivable that this considered view of utilitarianism might be forced to endorse the "removal" of the minority in the example presented above. Indeed, even if the circumstances are such that the removal would *not* produce the greatest amount of satisfaction when compared with the alternatives, Williams plainly thinks there is something deeply unsettling about the satisfaction of the majority counting as part of of the desirable state of affairs *at all*[35]. Second, this revised view fails to make sense of the difficulty in Jim and George's cases:

> The decision so determined is, for utilitarianism, the right decision. But what if it conflicts with some project of mine? This, the utilitarian will say, has already been dealt with: the satisfaction to you of fulfilling your project, and any satisfactions to others of your so doing, have already been through the calculating device and have been found inadequate. Now in the case of many sorts of projects, that is a perfectly reasonable sort of answer. *But in the case of projects of the sort I have called 'commitments',* those with which one is more deeply and extensively involved and identified, this cannot just by itself be an adequate answer, and there may be no adequate answer at all. For, to take the extreme sort of case, how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because someone else's projects have so structured the causal scene that that is how the utilitarian sum comes out?[36]

[35] Ibid, 105.
[36] Ibid, 115-6. Emphasis added.

Jim's misgivings about killing an innocent prisoner *are* relevant to the utilitarian decision in that case, but "their weight must be small: they are after all (and at best) one man's feelings"[37]. And for another agent, this might seem reasonable. Looking down at Jim as he makes his decision, his commitments seem trivial against the lives of the nineteen, and the related suffering he would bring to their families. But Williams' argument is not that all agents must develop some loyalty to everyone else's projects, nor that one's commitments will always sustain conclusive reasons when a conflict arises. Williams hopes to make a claim about the relation between an agent and their own values, between individuals and the projects that they have. As we have seen so far, commitments are more than inclinations against actions that strike an agent as unpleasant, they are representative of the agent's ethical identity, of their perspective on value. And so when Williams asks "how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life", he is not suggesting that it would be *unfair* to expect an agent to abandon a commitment because they have *invested* too much of their life into it. Instead, he is claiming that this is something a rational agent *simply cannot do*. Perhaps, despite his commitment, Jim can make the difficult choice to kill the prisoner, but he cannot understand the circumstances as obvious unless he is able to view the circumstances from the outside, as though he is not the same agent with those commitments. He *is* himself, and so the choice he is forced to make cannot be made easily. If it were, and Jim could view his commitment as only instrumental to some greater value, then this would only be proof that he does not have that commitment at all:

> …once he is prepared to look at it like that, the argument in any serious case is over anyway. The point is that he is identified with his actions as flowing from projects and attitudes which in some cases he takes seriously at the deepest level, as what his life is about… It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step

[37] Ibid, 102.

> aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which *his* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity.[38]

Any agent who is not a throughgoing utilitarian will have commitments which are not consistent with utilitarianism, and so might well find themselves in an ethical dilemma, even if a utilitarian solution is easily grasped and acknowledged. What's more, utilitarians must include amongst the outcomes maximized agents enjoying projects which they find intrinsically valuable. The trouble is, this will include commitments which conflict with utilitarian calculus, as having a preferential care for the welfare of others is often the most crucial sort of project available in human life, and one which is inconsistent with impartial satisfaction maximization. And so, it is impossible for the philosophy to have it both ways: to both cultivate and promote the projects most important in the pursuit of satisfaction in the patients of the moral calculus, and then to demand that those very individuals deliberate as though they never had those projects in the first place. It would be to declare that one project—the utilitarian project—has the authority over how an agent is allowed to view the value of their own commitments in deliberation, when utilitarianism *requires* that an agent values those commitments intrinsically. It would be to ignore the relationship between an agent's character, values, and their ability to choose. It would be to attack the very idea of their integrity as an individual person, capable of both action and feeling.

5. What is Integrity?

The formulation of the integrity objection from 1973's "A Critique of Utilitarianism", argues that there are certain important but neglected facts about agency, namely that agents will have commitments which threaten to conflict, placing the agent in an ethical dilemma. So, the

[38] Ibid, 116-7.

utilitarian attitude regarding the cases of Jim and George—that their choice should be obvious because of the clear difference in utility between their options—is inappropriate, as it assumes that Jim and George are throughgoing utilitarians, whose only commitment is to utilitarianism. But most people will not be throughgoing utilitarians, and so to presume that utilitarianism, which is at best one commitment among many, will have such authority over the agent so as to command them to disregard their other commitments is absurd—cheifly because it is impossible for the agent to see their circumstances as though they were not themselves: a definite character with a defined point of view. That they *are* themselves is identical to saying that they have integrity. They have a character, constituted by a suite of dispositions, including their projects, inclinations, commitments, desires, etc. This fact (which Williams takes considerable pains to show is not only independently plausible, *but necessary for a plausible utilitarianism*) animates the tension at the heart of the argument. That tension is between the fact that an agent will have contingent ethical motivations attached to their beliefs, values, and personal relationships, and the claim of a moral theory that all of their ethical reasons can only come from that impartial moral theory.

In my view, much of the ground covered in "A Critique of Utilitarianism" is ancillary to the integrity objection, pressing this tension onto utilitarianism in particular. As such, I think that much of what he has to say in that paper is in service of a secondary aim, which is to show that utilitarianism assumes a view of agency which requires integrity to make sense of the way the principle of utility can find determinant ways to promote satisfaction. But Williams' interest in integrity predates his application of it to utilitarianism in particular. In the 1965 essay "Ethical Consistency", Williams makes largely the same argument in a more general form, one which distinguishes what we have called commitments from other projects, and argues that when an

agent must choose between their commitments, their choice will not be easy, nor obvious in the way moral philosophers often require. Since it presses the same central points, I argue that the 1965 argument follows close enough to the 1973 argument that they should be understood as two instances of making the same objection, only that "A Critique of Utilitarianism" is augmented to stress the particular ways utilitarianism ignores integrity.

"Ethical Consistency" aims at comparing conflict between an agent's beliefs, their desires, and their ethical judgements[39]. Williams' method seeks to determine whether a suitable policy of resolving conflict between two or more ethical judgements better resembles a fitting policy for resolving conflicts between desires, or conflicts between beliefs[40]. While conflict between beliefs can be avoided through the suspension of judgment, and an agent can attempt to limit conflict between desires by discouraging one desire or the other,[41] an agent cannot freely choose to discourage an ethical judgement within themself, without only revealing that the agent was not attached to that judgment in the first place. Williams writes:

> A man who retreats from moral conflict to moral indifference, however, cannot at the same time admit that those conflicts were what, at the time, he took them to be, viz. conflicts of moral claims, since to admit that there exist moral claims in situations of that sort is incompatible with moral indifference towards those situations… A moral observer cannot regard another agent as free to restructure his moral outlook so as to withdraw moral involvement from the situations that produce conflict; and the agent himself cannot try such a policy, either, so long as he regards the conflicts he has experienced as conflicts with a genuine moral basis.[42]

For Williams, these ethical judgements are stringent in a way which desires are not, so that we cannot simply choose to discourage them within ourselves. Insofar as they are ethical claims attached to an agent, we actually cannot regard them as not worth pursuing or promoting.

---

[39] In the original paper, Williams refers to "moral judgements". Here, I will employ the term "ethical judgement", to signify that they need not be connected to a moral system.
[40] Williams, "Ethical Consistency", 166
[41] Ibid, 177-8.
[42] Ibid, 178-9.

So what is the nature of these ethical judgements, and what does it look like when they are in conflict? Williams focuses on cases where two unrelated ethical judgements happen to be in conflict due to some contingent circumstances. So, rather than discussing an agent whose ethical judgments are intrinsically inconsistent (such as those which might appear as the thoughts "I ought not be a thief" and "I ought to steal from the rich"), he offers cases where an agent's ethical judgments conflict by happenstance. His primary example is that of Agamemnon, whose ethical judgements are involved in his roles as a commander of a fleet of ships, and as the father of Iphigenia.

Agamemnon

> Agamemnon has offended the Goddess Artemis, and so she prevents his fleet from sailing to war in Troy, unless he offers the Goddess a precious sacrifice: the life of his daughter. As the commander, he is bound to arrive at Troy with his fleet, as well as protect the lives of his men. But as a father, he is bound to the welfare of his daughter.

The drama of this case comes from the conflict between Agamemnon's ethical judgements. It is natural to Agamemnon to affirm both of the following: "I must get the fleet to Troy safely", and "I must protect the welfare of my daughter", only he is now forced to act on one by neglecting the other. He eventually chooses to sacrifice Iphigenia, but Williams' point is that if he judged that he was ethically bound to the welfare of his daughter, he will be wracked with regret nevertheless. Regret takes on a technical meaning here, which does not require Agamemnon to later come to the conclusion that he should have chosen otherwise. He writes:

> The agonies that a man will experience after acting in full consciousness of such a situation are not to be traced to a persistent doubt that he may not have chosen the better thing; but, for instance, to a clear conviction that he has not done the better thing because there was no better thing to be done.[43]

[43] Ibid, 173.

When Agamemnon chooses his fleet over his daughter, his judgement regarding how he should treat Iphigenia is not revised to now appear to him to be not normative (as might happen when a conflict of beliefs is resolved, and one belief is abandoned when the agent becomes convinced of the truth of a conflicting belief). That is, it is not as if he assumed some *all-things-considered* deliberative position, which concluded that since acting on one ethical judgment would obstruct the other, that the losing ethical judgement was not *binding* in this particular case. His obligation to his daughter remains as strong as it did before, indeed it is inescapable. Agamemnon will see himself as having neglected one of his own values, even if he could not have acted in a way that satisfies all of his commitments in this case.

This idea should be familiar. Jim has a commitment, which motives him to refrain from killing innocent people, one which makes his choice difficult. But from the utilitarian perspective, everything from the lives of the prisoners, the welfare of their families, the projects of the warlord, Jim's misgivings, the political effects of the executions, and the possible logistical ramifications of saving bullets, all return to the moral calculus under the guise of *utility maximization*, the universal value through which all things are judged. And so, the utilitarian needs only acquire the facts, run the numbers, and voila! deliberation reaches its satisfaction. But Jim's commitments prevent him from seeing the killings as fungible tokens to be accounted for, and so he is unable to whole-heartedly submit to the calculation. As he is not a thoroughgoing utilitarian, the choice cannot be so easy for him. He may be a partial utilitarian, though, and grasp that from an external perspective, shooting the prisoner is the right thing to do[44]. But the effect of his having other commitments in conflict makes his circumstances dilemmic, that is why the choice cannot be made so easy.

[44] As Williams himself supposes, "A Critique of Utilitarianism", 117. It is important to see that his argument does not conclude in the objective value of the objects of commitments or ethical judgments, but rather a moral-psychological point about what deliberation looks like with real individuated agents.

Agamemnon might reach some deliberative conclusion, but that conclusion will not be satisfactory insofar as his ethical judgements are in conflict. He has a *commitment* to his daughter, which prevents him from regarding his daughter's life as a fungible value which can be exchanged for the welfare of his fleet. The answer is anything but obvious, and any choice will leave him with regret.

"Ethical Consistency" presses the same objection as "A Critique of Utilitarianism", that an agent's ethical point of view will often lead them to consider certain features of a case as ethically salient, features which cannot register in the deliberative schemes of impartial moral systems. The tension in the integrity objection is between the fact that an agent with integrity will be motivated by the commitments they have, and the assumption that impartial moral systems offer a decisive and emotionally satisfactory mode of deliberation. In the later paper, Williams also argues that utilitarianism requires agents to have projects and specifically commitments which will lead them to have misgivings about impartial deliberation. But that is not the essential point about integrity, only the form the argument took in a polemical essay directed at attacking utilitarianism. Indeed, even if utilitarianism did not require those assumptions, Williams' objection would work largely the same, because he still holds that agents will naturally develop commitments, that they are crucial for a satisfactorily meaningful life, and that their presence can threaten to make deliberation dilemmic.

6. Conclusion

The integrity objection works by appealing to what he considers a basic feature of ethical deliberation, that agents have commitments which can manifest as emotional dispositions for or against some acts. For most people, some things strike us as important and worth doing, taking the form of personal loyalties, aversions to some forms of conduct, or explicit moral precepts,

etc. Williams takes for granted the fact that nearly everyone will have such commitments, and claims that they are important for the pursuit of a meaningful life. That an agent's deliberation involves a determinate set of personal projects and commitments is what Williams calls integrity.

The objection works by demonstrating how in some cases of ethical dilemma, impartial morality systems neglect integrity in disquieting ways. In the case of Jim and George, utilitarianism cannot recognise the fact that as agents with integrity, and therefore commitments, and Jim and George cannot regard their own cases as obvious. And in the case of Agamemnon, his commitments make it so that his case cannot be obvious, no matter how it is resolved. And we sympathize with his protagonists, Williams hopes, because we are not thoroughgoing impartial agents, and we recognise that their cases will not feel obvious or simple. So, moral theory seems to lose out on an important piece of ethical life, insofar as it cannot account for integrity.

Chapter Two: Integrity and Moral Demands

1. Introduction

The integrity objection is counted amongst the most potent and enduring objections to consequentialism[45], though the force of the objection is impeded by obscurities in interpretation. Indeed, in the canonical formulation of the objection, the feature of consequentialism to which Williams objects is obscured by his aims in the polemical text *Utilitarianism : For and Against*. Here, I will make the case that the feature Williams finds objectionable has little to do with consequentialism itself, but rather with one's attitude towards the prescribed deliberation of any impartial moral theory. I hope to show that as a result, attempts to evade the objection by revising the rulings of a moral system are hopeless.

At the heart of the objection is an appeal to "absurdity"; Williams claims there is something absurd about the way that impartial morality systems infringe on integrity. Samuel Scheffler, in his *The Rejection of Consequentialism*, interprets this absurdity as stemming from the injunction to act in such a way that constitutes abandoning one's most seriously held commitments or relationships. To accommodate this interpretation of the objection, Scheffler breaks with consequentialism, and endorses when he calls an *agent-centered prerogative*, a normative consideration that permits agents to view their own commitments out of proportion with the way in which they are represented impartially by consequentialism. Here, I will argue that Scheffler misinterprets what is supposed to be absurd in Williams' cases, and so his accommodation fails to meet the problem as Williams proposed it. Namely, Scheffler is mistaken to take the problem as relating to the extreme demands of consequentialism. Instead, the

[45] For economy's sake, I will render this discussion in terms of consequentialism, though at times the authors mentioned concern themselves with utilitarianism. An important upshot of my argument is that the integrity objection in fact cannot be evaded by any structural device which could constitute a distinction between consequentialism and utilitarianism, anyways.

objection is about the way that impartial morality systems instruct agents to deliberate as though they are not themselves, and treat their seriously-held projects as though they are illegitimate sources of reasons. Indeed, the integrity objection is not about consequentialism in particular, it is about the way agents are permitted to relate to their own values. Or so I will argue.

2. The Integrity Objection

There are many well-worn objections to consequentialism which take the following form: 'No plausible normative system will ever instruct an agent to ϕ, and in some circumstances, normative system X will instruct agents to ϕ, therefore normative system X is an implausible normative theory'. The integrity objection is not an objection of this kind. Rather than evaluating the conclusions the normative theory is likely to recommend, the integrity objection questions the method of practical deliberation presupposed by consequentialism.

In this section, I will make use of the interpretation of the integrity objection set out in Chapter One, in which an agent's commitments prevent them from taking up the impartial point of view. It is in the nature of commitments that they loom large amongst an agent's deliberative landscape, appearing stringent, and compelling. Consequentialism ignores this fact. Consequentialism as a normative theory instructs agents how to rank-order states of affairs based on the relative prevalence of some desired consequence, *viewed impartially*. It then instructs agents to act in the way which produces the best ranked state of affairs possible. But since the states of affairs are ranked impartially, one's own projects and relationships are only represented insofar as they contribute to some impartially desirable consequence. The problem with consequentialism, for Williams', is *not* that it might instruct an agent to do something against their own interests (as any suitable moral theory might), but that it instructs agents to deliberate

through a view of the world which is irrational to occupy, where one's commitments are treated with apathy.

A crucial point is that this is an objection from *procedure*; the objection stands regardless of whether the action ultimately instructed by consequentialism differs from the action endorsed by an agent's commitments. I have claimed that besides the famous formulation in "A Critique of Utilitarianism", Williams makes the same objection in an earlier paper "Ethical Consistency". This is because the earlier paper argues that an agent's ethical judgements have a motivational independence, such that they will offer reasons for action to the agent who has them, and move them to regret their neglect even when they reach an *all-things-considered* decision to act against them. I have claimed that these ethical judgements are what Williams will later call commitments. In another essay, entitled "Persons, Character and Morality", Williams again mobilizes commitments against the mode of deliberation proposed by some moral theories. The relevant part of the paper makes use of a case involving a 'rescuer', making a point about what sorts of motivations should be available to beneficent agents.

Rescuer

> A natural disaster forces a private citizen to engage in heroics. Looking to aid some victim of the disaster, they push into a building and search an open room. Inside, they find two individuals: a stranger, and their own spouse. Knowing they only have time to save one, the rescuer saves their spouse[46].

In 'rescuer', Williams takes aim at moral systems which instruct agents to deliberate only on the basis of general principles. He makes this point by conjecturing about the reasons that might have motivated the rescuer to save their spouse. Some philosophers might attempt to justify the

[46] Bernard Williams, "Persons, Character and Morality" in *Moral Luck : Philosophical Papers, 1973-1980*, (Cambridge University Press, 1981), 17.

rescuer's decision by appealing to a general moral principle. They might want the rescuer to think: "it was [my spouse] and that in situations of this kind it is permissible to save one's [spouse]". But Williams thinks this is "one thought too many". If the rescuer has a genuine commitment to their spouse, then the motivating thought will not involve some external appeal to impartial permissions; they need only think 'it was my spouse'[47]. Or at least, this reasoning ought to be available to the rescuer.

The problem here is not that a moral system delivers an unsatisfactory answer. Rather, the problem is that the demand that they reason from general principles misunderstands what it would be like for the rescuer to deliberate over this case. One might rationally accept consequentialism, but when the rescuer enters the room with their spouse and the stranger, it seems unsuitable to hold them responsible for partiality. This is the meaning of integrity; the rescuer's ownership of their commitments and relationships are deeply connected to who they are as an agent, and so to recast practical deliberation in exclusively impartial terms cleaves them from their own point of view.

Notice that in this case, whether or not consequentialism actually instructs an agent to abandon their projects *in action* is immaterial. Indeed, in 'rescuer' the moral theory may endorse the agent's act, just as in 'Jim', Williams suggests that the consequentialist might even be right about what Jim ought to do[48]. But this is a procedural objection, in that what is at issue is how moral decisions are made, and not the content of the decisions themselves. The objection is that it is rational that we will weigh our commitments out of proportion with their impartial worth under consequentialism, and that consequentialism fails to accommodate this fact.

[47] Ibid, 18.
[48] Williams, "A Critique of Utilitarianism", 117.

3. The Agent-Centered Prerogative[49]

The Rejection of Consequentialism is presented as a systematic answer to the integrity objection, which works to accommodate the worry whilst retaining some of the intuitive appeal of consequentialism. Scheffler begins by characterizing the integrity objection as responding to a systematic "[severing] between one's deepest commitments and concerns… and one's actions"[50]. Scheffler takes Williams' objection to regard some disconnect between an agent's ability to act on their seriously held values (for instance, Jim's seriously-held value of non-violence) and the way in which consequentialism makes it seem all too obvious that Jim ought to kill the prisoner.

Scheffler claims that the objection is 'suggestive', but 'obscure'. The obscurity comes from Williams' failure to indicate exactly which feature of consequentialism threatens to undermine integrity, as Williams cannot mean that an agent has a claim against consequentialism whenever the theory instructs against *any* of their projects. If Williams meant that it must be permissible to pursue *any* interest at *any* juncture, then the only moral theory which could satisfy his objection is egoism. He writes:

> But if it is this feature of utilitarianism that attacks the agent's integrity, it is doubtful that any theory but complete egoism could avoid doing so. For virtually any moral theory will make the permissibility of pursuing one's own projects depend at least in part on the state of the world from an impersonal standpoint. Virtually any moral view will hold that if things get bad enough from the impersonal standpoint, the agent's projects become dispensable.[51]

But Scheffler does not think that the integrity objection is meant to object to all moral theories which might, under any circumstances, instruct an agent to abandon one of their projects. He believes the objection is about consequentialism, and so he reconstructs the feature of consequentialism which alienates integrity:

---

[49] This section is exegetical, focusing on Scheffler's interpretation of the integrity objection. Scheffler will not use "moral" or "ethical" in a way which falls in line with Williams' moral/ethical distinction.
[50] Samuel Scheffler, *The Rejection of Consequentialism : A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, (Clarendon Press, 1994), 7.
[51] Ibid, 8.

> [The integrity objection] should be seen as arising… in response to the discrepancy between the way in which concerns and commitments are *naturally* generated from a person's point of view quite independently of the weight of those concerns in an impersonal ranking… and the way in which [consequentialism] requires the agent to treat the concerns generated from his point of view as altogether dependant for their *moral* significance on their weight in such ranking. In other words, [consequentialism] incorporates a conception of the right which requires each agent in all cases to produce the best available outcome overall.[52]

Scheffler emphasizes the failure of consequentialism to reflect the natural way in which people are inclined towards the interests they are partial to. It prohibits them from "treating" their interests in accordance with this personal value, and excludes any partial interest from the obligatory conception of right, except insofar as that interest is represented impartially.

For Scheffler, an attractive moral theory ought to reflect the way in which people naturally engage with value. He accomplishes this end, by offering a "hybrid" normative theory, which is largely consequentialist, but departs with what Scheffler calls an *agent-centered prerogative*, which amounts to permission for agents to treat their own projects and interests out of proportion with their value when considered impartially[53]. Of course this is quite vague, though Scheffler only hopes to show that it is reasonable to hold that *some sort* of prerogative exists, without committing to the amount of leeway it lends agents in fact[54]. Though, Scheffler clearly thinks that some versions of this prerogative are more tenable than others. He starts by restating his motivation:

> If the unrestricted responsibility for producing optimal outcomes that consequentialism assigns to individuals is thought to be objectionably demanding, then the natural solution is to allow agents not to promote such outcomes when it would be unduly costly or burdensome to do so.[55]

This 'natural solution' must then define 'objectionally demanding'. Scheffler quickly rejects an unlimited prerogative, which would allow the agent total freedom to treat their interests with

[52] Ibid, 9. Scheffler's emphasis.
[53] Ibid, 14.
[54] Ibid, 69.
[55] Ibid, 20.

inflated weight however they will, seeing this as threatening a collapse into egoism[56]. Of course, Scheffler had already rejected this approach in his reconstruction of Williams, when he interpreted Williams as viewing consequentialism as threatening integrity insofar as it undermined *some* of the agent's crucial projects. As we saw, if we interpret Williams as thinking integrity is threatened whenever *any* agent-relative project is threatened, then only an egoism could satisfy the animating intuition, says Scheffler. He concludes that this is therefore an untenable interpretation of Williams, and subsequently offers the reconstruction above.

Scheffler then makes a surprising move. In comparing the relative merits of a consequentialist attempt to evade the integrity objection with his own hybrid model, Scheffler does not determine whether or not consequentialism is capable of defeating the objection without the modification he suggests[57]. This is surprising, as to this point he has offered the agent-centered prerogative as an answer to the integrity objection. However, it is Scheffler's position that there exists a deeper foundation for the integrity objection which has not yet been made clear, and which motivates an independent rational justification for the agent-centered prerogative[58]. This foundation is a deep-seated uneasiness, which is connected to an inability for consequentialism to accommodate "the independence of the personal point of view"[59]. He articulates the intuition:

> From the impersonal standpoint, the fact that some person has a special kind of concern for his own projects and plans is relevant only in so far as it may affect the size of the harm or benefit incurred by that person when those projects fare poorly or well. Once that harm or benefit, however small or great, has been figured into the overall evaluation, the impersonal standpoint has no *further* concern for his projects. But *his* point of view is independent of the impersonal point of view, and this independence is typically

[56] Ibid, 21.
[57] Ibid, 53.
[58] Ibid, 54.
[59] Ibid, 56.

> evidenced by the fact that *he* cares differentially about his projects just because they are *his* projects.[60]

This, Scheffler says, is the intuition which motivates Williams' integrity objection, and he turns his intention to answer this intuition, rather than the integrity objection as construed above. The intuition is that moral theory must somehow accommodate an agent's personal perspective and values in this evaluation of their projects, and that consequentialism, which treats each person's project equally, cannot do this.

Scheffler claims that a hybrid theory that includes the agent-centered prerogative can meet this deeper objection. He offers two reasons in support of this claim. The first reason is that Scheffler sees it as important that a moral theory incorporates the natural independence of the personal point of view into its deliberation; a normative system that obligates an agent to occupy the impersonal point of view is unnatural, as "actions [do not] typically flow from the kinds of concerns a being who actually did inhabit the impersonal standpoint might have"[61]. And so a normative system must somehow accommodate this typical relation between actions and motives. Moreover, this fact must be accommodated in the right way. One strategy a consequentialist might deploy to make this accommodation is to include the pursuit of projects for all in one's conception of the good, a good which is then maximized following an impartial prescription[62]. However, Scheffler's second reason in support of the agent-centered prerogative works to prove that the hybrid model is better at *appropriately* accommodating the way personal interests naturally affect deliberation. He writes: "a moral view gives sufficient weight to that fact only if it *reflects* it, by freeing people from the demand that their actions and motives always

[60] Ibid, 56-7. Scheffler's emphasis.
[61] Ibid, 62.
[62] Ibid, 60.

be optimal from the impersonal perspective, and by allowing them to devote attention to their projects and concerns to a greater extent than impersonal optimality by itself would allow"[63].

According to Scheffler, the incorporation of an agent-centered prerogative into a normative theory makes it a preferable alternative to consequentialism–even a consequentialism tilted towards supporting the pursuit of the projects of agents. This is because the hybrid theory better reflects the way in which an agent interacts with their own projects and interests, a normative reflection of psychological fact. This amounts to a rejection of consequentialism, as it is no longer the case that the normative theory instructs an agent to deliberate to the exclusive end of maximizing desirable consequences in their actions. Rather, the normative theory permits the intrusion of personal commitments into deliberation, without demanding that the agent always act in line with the consequentialist project of maximization.

4. The Meaning of Integrity

Scheffler portrays the integrity objection as an uneasiness with the consequentialist injunction to always act in such a way as to maximize desirable consequences. Scheffler reconstructs the integrity objection as an objection to the absurdity of demanding that an agent acts in accordance with utility "when it would be unduly costly or burdensome to do so"[64]. Indeed, when motivating the integrity objection, it is instructive that Scheffler always seems to be concerned with breaking with one's projects *in action*, rather than *in deliberation*–with what one is commanded to do, rather than how one views their interests. But Williams' integrity objection is an objection about deliberative procedure, not about demanding conclusions. This is most evident in two cases offered to illustrate the integrity objection, 'Jim' and 'Rescuer', which do not turn on what an agent does, but on how the agents are expected to determine what they

[63] Ibid, 62.
[64] Ibid, 20.

do. Williams is instead concerned with how an agent is supposed to occupy an impartial perspective *at all*, whilst under the influence of the commitments they have and value. Williams: "the point is that somewhere… one reaches the necessity that such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view"[65]. In this argument, Williams is not unsatisfied with the demanding character of the general principles of consequentialism, he is skeptical that one can rationally make sense of their personal commitments in terms intelligible as general principles at all. It seems, commitments from the personal perspective cannot be satisfied by *any* impartial moral theory, as it undermines the integrity of the values of the agent to seek moral permission for expressions of that value. Actions from moral principles are expressive of the doctrine of morality, and that is "one thought too many" to be an expression of one's own integrity.

Scheffler errs when he reconstructs the integrity objection in terms of action. Doing so misses the procedural worry that Williams is trying to express, and leaves one puzzled as to how the cases Williams offers are good examples of the objection Scheffler construes (if in both cases personal commitments conform with morality *in action*, where is the demandingness?). There is another problem with Scheffler's reconstruction, which is brought out by his attempt to exclude a seemingly implausible interpretation of the integrity objection. As noted above, Scheffler worries that if the integrity objection relies on the "in-principle dispensability of the agent's projects", then virtually the only theory which evades this charge is egoism[66]. This cannot be, and since Williams raises this objection with consequentialism in mind, and so it must be that the charge has to do with some distinctive feature of consequentialism[67]. Later, this led Scheffler to reject an unlimited prerogative, where an agent would have permission to refuse the instructions of

[65] Williams, "Persons, Character and Morality", 18.
[66] Scheffler, *The Rejection of Consequentialism*, 8.
[67] Ibid, 9.

morality whenever it conflicts with their projects and interests. Scheffler abandons the interpretation of Williams that could only be satisfied by this unlimited prerogative, but in doing so placed a condition on the sorts of projects which could be protected by the agent-centered prerogative. Characteristically, Scheffler does not commit to any particular version of the agent-centered prerogative, and so we do not learn the nature of this condition, though it is notable that when he compares the relative plausibility of different versions, he does so on the basis of the relative *demandingness* of the suggestions[68]. Regardless, there is some condition placed on which projects can be protected by the prerogative, which Scheffler believes resists the objection by revising the demandingness of consequentialism. And so, the agent-centered prerogative extended to agents is subject to moral limitations.

Scheffler's reconstruction of the integrity objection is in part motivated by attempting to discover the distinctive feature of consequentialism which poses a threat to integrity. He suggests that this distinctive feature is consequentialism's *strong doctrine of negative responsibility*[69], which makes an agent responsible for their failure to prevent objectionable outcomes, even if those outcomes are unrelated to their own projects and actions. Indeed, an agent would be instructed to abandon their personal projects, even to prevent negative outcomes produced by the actions of others. This is one of the points brought out by 'Jim' above; the permissibility of one's projects is at the mercy of the activity of others. Though while this can be taken as a uniquely potent threat to integrity, Williams does not mean to say that this is the feature which makes consequentialism vulnerable to the integrity objection in the first place. Indeed, the integrity objection is *not* an objection to consequentialism, but to impartial moral systems in general.

[68] Ibid 20, footnote 6.
[69] Ibid, 7.

Williams develops the objection originally made against consequentialism to one which applies to moral theory in general:

> A man who has [a seriously valued] project will be required by [consequentialism] to give up what it requires in a given case just if that conflicts with what he is required to do as an impartial utility-maximizer when all the causally relevant considerations are in. That is quite an absurd requirement. But the Kantian, who can do rather better than that, still cannot do well enough. [...] There can come a point at which it is quite unreasonable for a man to give up, in the name of impartial good… something which is a condition of his having any interest in being around in that world at all.[70]

The argument is not contingent on the features of consequentialism, or even Kantianism. Instead, the distinctive feature from which the integrity objection springs is the attempt by moral philosophy to legislate over how one views their own values. Scheffler's reconstruction fails, because it is articulated to meet an instantiation of the objection which is suited to emphasize the peculiar and strikingly demanding way in which consequentialism threatens integrity, but loses sight of the deeper point about agency and impartiality.

Scheffler gets closer with the deeper intuition he thinks underlies the integrity objection, but he does not get close enough. This deeper intuition gestures at a concern for individual personhood which is indeed at issue in the integrity objection, however its complaint is that moral theory doesn't do enough to *reflect* the natural way in which individuals relate to their projects and interests. This leads Scheffler to propose a new moral theory, one in which morality can endorse an agent in invoking their prerogative to treat their interests with concern out of proportion with the impartial value of those interests. Indeed, it offers permission to agents to do so. It provides a device *within* a normative theory to allow agents to devote attention to their projects and relationships, a freedom from maximization which is subject to moral limitations.

But this is not an adequate answer to the integrity objection. As I have striven to emphasize, the reasonableness of the actions instructed by a moral theory are not at issue here,

[70] Williams, "Persons, Character and Morality", 14.

and so by softening the demandingness of consequentialism, Scheffler addresses the way in which consequentialism is only *primus inter pares* insofar as moral theories systematically undermine integrity. But the integrity objection is not about consequentialism after all; it does not arise from the commands of any moral theory, but from the assumed stance an agent is expected to take in response to morality.

Consider 'rescuer'. In this case, the hybrid theorist would have them thinking "it was [my spouse] and that in situations of this kind it is permissible to save one's [spouse]"? In this case, the decision is only settled with a thought for moral permissibility from general principles. This is one thought too many, Williams would say. They are not moved to saving their spouse by the rational authority of their commitment, but rather granted permission to chase a desire, on the authority of the hybrid theory.

The same is true in 'Jim'. When Jim is offered the infamous deal, the conflict he faces persists if the agent-centered prerogative does not determine that the prospect of his going through with the killing is burdensome enough. And if it is so flexible, then Jim is still depending on an impartial perspective for permission to pursue a project that is only intelligible from *his* perspective. Indeed, this is the case because the agent-centered prerogative attempt to "naturally" accommodate the personal perspective of agents grants them an amount of leeway which is externally mandated. Remember, for Scheffler the prerogative is limited to only grant an agent permission when it would be sufficiently burdensome not to, and so whether an agent recipes that permission is subject to the external evaluation of the effect of the value. But a commitment, for Williams, does not require this external evaluation to offer its agent a reason for action; it simply is reason bearing because it is a motive of the agent.

In George's case it is the same. If the agent-centered prerogative will not grant George permission to act on the reasons which came from his commitment, is George expected to conclude that he has *no reason* that comes from that commitment? And if he is granted that permission, that would be the case based on how burdensome denying him would be. But does George's case seem so constructed to suggest that it would be burdensome to take the job? After all, he would gain numerous personal benefits, including receiving the money to support his family. But taking the job would weigh on his conscience, and violate his commitment. This is a distinction that Scheffler cannot register; George is motivated by a commitment, not a mere desire. Surely, setting one's ordinary desires aside is burdensome, but it misunderstands Williams' objection to accommodate it with a deliberative device which treats commitments the same as it does other desires.

Considered in terms of the hybrid theory, these agents are still assumed to be required (in some sense) to resolve themselves to action with reference to morality, for permission in deliberation concerning their most serious commitments. But since that permission is external to themself, and impartial, it is incapable of making sense of those commitments, the only perspective capable of doing justice to the value of one's projects is one's own.

In its 1973 formulation, the integrity objection targets utilitarianism in particular, however the upshot of the objection implicates the decision procedure of any moral theory that demands impartial deliberation. When an agent has a commitment, they will not be able to deliberate as though that commitment does not exist, and so that it is not a legitimate source of reasons for them. While it might so happen that the principle of utility, or the categorical imperative, endorses some action which happens to promote or project a value involved in some agent's commitment, serving that commitment on its own merit and on the authority of impartial

morality are very different motivations. And the more personal motivation will appeal to agents with commitments. Williams calls the attachment an agent has to their commitments (and other motives) their integrity, and the objection is not about particular instances where an agent acts against those commitments, but about any moral theory that fails to respect the integrity of agents.

5. Conclusion

As I read the integrity objection, the problem with impartial morality systems is not that they are too demanding, but that they can only make sense of one source of demands. Williams denies that an agent's ethical reasons can only come from that impartial moral system. Scheffler's suggestion—the agent centered prerogative—attempts to resolve the integrity objection by making ethics less demanding, by giving agents permission to sometimes opt out of the injunctions of consequentialism if those injunctions are too demanding. But this misses the point. Williams does not want Jim to be released from the burden of morality, he only emphasizes that he is subjected to the competing reasons of his commitments, and so it is not obvious that he is rationally required to do the utilitarian thing. And in rescuing their spouse for partial reasons, the rescuer is not shirking their responsibilities, but living up to them—it is only that this responsibility cannot be represented without reference to the partial, first-personal commitments contingent on their own motives.

Scheffler interprets the integrity objection as a form of the demandingness objection, taking it to complain on behalf of agents who are forced by impartial moral systems to act in a way which destroys their personal projects and interests. The worry is that the demands of consequentialism are unmanageable, and so consequentialism must be replaced with some moral system which can grant the contrition dearly needed by agents squeezed out of their own projects

and lives by morality. This new moral system resembles consequentialism, though with the addition of the agent-centered prerogative, which gives agents permission to act in the pursuit of their partial, personal interests over the interests of morality, only when morality is exceptionally demanding. But the integrity objection is not about burdens, it is about authority. William's worry is not that it might be too hard for Jim or the rescuer to act impartially, it is that they will not *think* impartially if they have commitments. It is only with this in mind that we can make sense of the cases Williams offers where moral theory simultaneously endorses actions consistent with an agent's commitments, and yet seems to offend against their integrity.

Chapter Three: Integrity and Reason

1. Introduction

Bernard Williams' integrity objection contends that the moral theories which command agents to deliberate from a position of impartiality fail to recognise important facts about autonomous engagement with value, and that this failure undermines the rational authority of their prescriptions. Chappell argues that these purported facts about autonomy and value are best understood as relying on Williams' *internal reasons thesis*; they are facts about what can explain an agent's motivation to act, and what cannot. According to the internal reasons thesis, an agent can *only* have reason to ϕ if they "could reach the conclusion that he should ϕ (or a conclusion to ϕ) by a sound deliberative route from the motivations that he has in his actual motivational set - that is, the set of his desires, evaluations, attitudes, projects, and so on."[71] However, to always act impartially is to never have the opportunity to pursue one's own interests, and so is inconsistent with the desires from which any act must be justified. There can be no sound deliberative route from A to B if B precludes the possibility of achieving A. For Chappell, the integrity objection suggests that—given the internal reasons thesis—there is no reason to act impartially, as it would undermine the pursuit of one's crucial projects. And so, the internal reasons thesis plays a central role in explaining why the demands of an omnibearing impartial morality are distinctively problematic. But, Chappell continues, it cannot achieve that result for the integrity objection, because there is ambiguity in what counts as an "already-existing" motive which an agent must reason from. Further, any available interpretation of an "already-existing" motive also cannot achieve the result the integrity objection requires: it cannot explain what is uniquely problematic about the demands of impartial morality.

[71] Williams, "Internal Reasons and the Obscurity of Blame", 35.

I argue that Chappell's treatment of the integrity objection is misguided, for two principle reasons. First, the internal reasons thesis is not ambiguous in the way Chappell claims. Indeed, the seeming ambiguities are illusory, as Williams quite clearly defines the relevant features of the thesis in a way which makes an agent's reasons responsive to their individual character, involving individual contingent motives, restrained by a rational commitment to a sound relationship between an agent's motives and actions. Second, Chappell mischaracterises the relationship between the internal reasons thesis and the integrity objection. Chappell argues that the integrity objection claims that agents can derive no reasons from a system of impartial morality, because there is no sound deliberative route from motives to actions which undermine those motives. But the integrity objection does not make the claim that impartial morality systems cannot offer reasons to agents, it only that those reasons (should they exist) are not decisive; a rational agent will have reasons of other kinds which might triumph in guiding an agent's deliberation. Unpacking why Chappell's reading is in error will deepen our understanding of what is at stake in the integrity objection, most importantly by distinguishing it from other related concepts Williams employs.

2. Chappell's Objections[72]

2.1 Overview

For Chappell, the integrity objection is a form of the demandingness objection, which works to show that the demands of an impartial moral system are distinctly problematic next to the other sorts of reasons an autonomous agent might have. Of course, that morality makes demands of agents is appropriate, but these demands can become problematic when their pressure regularly impedes agent's ability to live their own lives: "The problem of

[72] This section is exegetical, focusing on Chappell's understanding of Williams' view. Chappell will use "moral" to describe a sort of motive which is "external" from an agent's already-existing desires. The term will otherwise resemble Williams' definition of the term, referring to the motive to act with impartiality.

demandingness is not about a theory's making extreme demands *sometimes*. The problem is that some moral theories seem to make extreme moral demands on us, not just sometimes, but always."[73] This prospect is unduly demanding, says Chappell, when morality endlessly commands substantial sacrifices, and exerts control over how we organise our lives at all times: "Such continual demands seem to threaten important parts of human life; in Bernard Williams' phrase, they threaten our 'integrity.'"[74] In Chappell's reading, integrity is our ability to exercise control over how we structure our lives. We are alienated from this ability when a system of impartial morality does not privilege our integrity enough to allow us to exercise that control. However, the relation between integrity and the demands of morality in Williams' objection is not straightforward, and indeed is often misunderstood due to the neglected role played by the *internal reasons thesis* in Williams' argument. All told, the internal reasons thesis should help explain why there is in fact *no reason* to follow the demands of impartial morality, because their threat to integrity shows that impartiality is incompatible with our motives. If this can be demonstrated, then the distinctive course taken by the integrity objection requires the internal reasons thesis to do the work of explaining why there is no reason to follow the demands of impartial morality.

But Chappell does not think Williams' integrity objection is sound. Chappell argues that the internal reasons thesis is ambiguous, and that any available interpretation of this ambiguity forces a reading of the objection which cannot do the work required in explaining how Williams' objection works. The ambiguity is in explaining what counts as an "already-existing motive", which according to the internal reasons thesis, is the necessary condition for there to be a reason for action. According to Chappell, any available interpretation of what counts as an

[73] Sophie Grace Chappell, "Integrity and Demandingness." *Ethical Theory and Moral Practice* 10, no. 3 (2007), 255.
[74] Chappell, "Integrity and Demandingness", 256

already-existing motive in the internal reasons thesis fails to support the integrity objection, as it represents the demands of morality as including trivial instrumental inferences which cannot be distinguished from the overbearing demands the argument means to highlight, or as already entrenched in a person's motives, failing to distinguish the demands from other, innocuous motives.

In this section, I review the internal reasons thesis. Then, I present Chappell's reconstruction of the integrity objection, including its relation to the internal reasons thesis. This reconstruction will include Chappell's critique of the objection, which suggests that there is a problematic ambiguity in the internal reasons thesis regarding what counts as an already-existing motive.

2.2 The Internal Reasons Thesis

The internal reasons thesis proliferates from one important claim: "An internal reason statement is falsified by the absence of some appropriate element from S", given that S refers to "the agent's *subjective motivational set*"[75] understood "in terms of desires… [but also] such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent"[76]. Given the internal interpretation, an agent has a reason for action only if that action will pursue the satisfaction of an element of S, which is composed of their various desires, projects and dispositions ("motives" hereafter). This claim is about the truth conditions for a statement about whether or not an agent has internal reasons, which provide an exhaustive list of the reasons for action an agent can follow rationally.

[75] Bernard Williams, "Internal and External Reasons" in *Moral Luck: Philosophical Papers, 1973-1980*, (Cambridge University Press, 1981), 102.
[76] Ibid, 105.

The motives which are elements of S inform an agent's reasons in one of two ways: (a) a member of their S is to undertake some action or (b) the agent, should they deliberate appropriately, can discover that they have reason to ϕ by discovering how ϕing serves a motive of theirs. This second option might admit of a further division, in that this deliberative process can take the form of (i) *instrumental reasoning* regarding how actions might satisfy members of S, either by obtaining some effect, or even only symbolically[77]. In such a case: "The internalist view of reasons for action is that… at least a necessary condition its being true that A has a reason to ϕ [is]: A has reason to do ϕ only if he could reach the conclusion to ϕ by a sound deliberative route from the motivations he already has."[78] I will call the second purported form of deliberation (ii) *persuasion*, an imaginative process through which an agent considers what the satisfaction of some motive might be like, moving them to more clearly understand what their motivational set entails, leading them to clarify the objectives of the motivations or perhaps to add or subtract a motivational element.[79] In practice, it would be hard to distinguish between (i) instrumental reasoning and (ii) persuasion, as externally it would be hard to say whether another's deliberation has found a reason for action based on a pre-existing motive, or if they have developed a new motive. Nevertheless, Williams offers this conceptual difference which supposes that there is some distinction between these processes.

This distinction is related to one feature of the internal reasons thesis, namely that an agent's reasons, relative to their motives as they are, are nevertheless not immediately transparent to an agent. By this I mean an agent may have reasons that they are not aware of, or be mistaken about the reasons they take themselves to have. These possibilities involve the

[77] Ibid, 103.
[78] Williams, "Internal Reasons and the Obscurity of Blame", 35.
[79] Williams, "Internal and External Reasons", 105.

relation between an agent's motives and reasons, loaded into what Williams calls a "sound deliberative route" from an agent's motivations to their reasons. Williams writes:

> An important part of the internalist account lies in the idea of there being a 'sound deliberative route' from the agent's existing S to his ϕ-ing. 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... One reason why it must do so is that it plays an important part in discussions about what people should become disposed to do.[80]

As we will see below, this commitment to soundness is a rational constraint on what an agent's reasons actually are, and so on the internalist view it is possible for an agent to be under the false impression that ϕ-ing follows from one of their motives, rendering their ϕ-ing irrational.

As I understand Williams, the internal reasons thesis is about what courses of actions, and grounds of those actions, are available to rational agents. For Williams, a rational agent will be responsive to the reasons they actually have, given their motives (the motives an agent has does not appear to be rationally constrained). An agent's reasons are determined given their motives, but they are not necessarily conscious of those reasons, and so to be genuinely responsive to one's reasons, an agent must engage in sound deliberation to determine what sorts of actions best serve their motives. When successful, the agent identifies their reasons for action, and if their reasoning was sound, they are rational by acting on those reasons.

2.3 Chappell on Integrity

For Chappell, morality characteristically requires the sacrifice of our desires; it purports to offer reasons to act against those desires. This can become problematic when morality demands that we must replace our existing motivations with impartial moral precepts, abandoning the dispositions which propel our personal lives[81]: "Such continual demands seem to

[80] Williams "Internal Reasons and the Obscurity of Blame", 36.
[81] Chappell, "Integrity and Demandingness", 256.

threaten important parts of human life[82]. Sensitivity to moral demands, then, is the sensation of a pressure external to the motives which are naturally a part of an agent's reasoning, and so the demandingness of impartial morality is then only intelligible when it takes the form of this external pressure on agents' conduct.

For Chappell, morality becomes "extremism" when it asks us to abandon our existing desires altogether, and replace them with only impartial moral rules. But morality systems which become extreme in this way are "pragmatically self-defeating", as "unless agents may have and be shaped by their own distinctive projects, 'dispositions,' and 'commitments,' they will have no reason to do *anything*; so in particular, no reason to do what extremism tells them to do"[83]. The idea, ultimately, is that a life lived without space to enjoy engagement with personal commitments will sap an agent's motivation to carry on with that life, and the agent will ultimately be left beleaguered and indisposed to the moral precepts which made their life so. For Chappell, such prospects are "pragmatically self-defeating" because there cannot be a *sound deliberative route* from one's desires to impartial morality, as a life of impartial conduct is destined to alienate an agent from those desires.[84] She writes: "this is the only way that reasons to respond to extreme moral demands could enter the scene: by my moving, from my most basic already-existing motives, along a sound deliberative route to the conclusion that I ought to do something about those demands; or at least about some of them"[85], but "no such deliberative route could be sound, if its end contradicted its beginning; as it would, if the end of the route was the conclusion that I ought to do things that would entirely rule out acting on my most basic already-existing motives"[86]. As Chappell understands moral demands, they are not naturally

[82] Ibid.
[83] Ibid.
[84] Ibid, 258.
[85] Ibid.
[86] Ibid.

members of S. The only way a purported reason to ϕ might become authoritative over an agent is if it can be found that ϕ-ing pursues satisfying an element of S, thereby identifying it as an internal reason[87]. But this process is constrained by the requirement that its reasoning is sound, and so it must really be the case that ϕing in fact does pursue satisfying an element of S, and not only appear to do so ahead of time[88]. Moral demands are too demanding when they violate this constraint, when following those demands makes the pursuit of one's own motives impossible. A life of perfect impartiality would force an agent to alienate their own motives forever, and so no form of sound reasoning could lead an agent to believe that following these moral demands is constitutive of, or even compatible with, pursuing the satisfaction of element of S. And so, there are no internal reasons to follow all of moralities' extreme demands[89].

Chappell's reconstruction can be characterized as relying on one key move. First, Chappell claims that the reasons offered by impartial morality are not rooted in elements of S, and so adhering to those reasons requires an agent to set their own motives aside. And so, the integrity objection explains the attack on integrity as moral reasons not satisfying the necessary conditions set out by the internal reasons thesis. The internal reasons thesis must determine what sort of already-existing motive can serve as an element of S (and what cannot), to explain what is problematic about moral demands in the integrity objection. Chappell will then go on to claim that the internal reasons thesis is too ambiguous to do this work, and any plausible mode of resolving this ambiguity remains unable to explain the problematic effect of impartial demands on integrity.

[87] Williams, "Internal and External Reasons", 109.
[88] Ibid, 102.
[89] Chappell, "Integrity and Demandingness", 258.

2.4 Already-Existing Motives, Chappell's View

With the internal reasons thesis playing such a crucial role in the integrity objection, the stage is set for Chappell to apply some pressure to the objection through the internal reasons thesis. According to Chappell the internal reasons thesis is ambiguous, to the point that it cannot specify what counts as an "already-existing motive"[90]. Further, the available plausible readings for what might actually count as an already-existing motive cannot explain why impartial morality is unduly demanding.

Chappell begins: "The notion of 'my already-existing motives' seems to admit equally well of an indefinite variety of narrower and wider readings."[91] As evidence, she offers two polar interpretations of the internal reasons thesis as valid readings of its central claims: "On the narrowest readings, [S] will, at any time *t*, include little more than those motives that I am consciously aware of at *t*", and "On the widest readings, [S] may have its contents determined by Thomist or Aristotelian claims about the essential nature of humans just as such, and about the sort of motivations that that essential nature makes universal in humans"[92]. She continues: "On such wide accounts of what 'my already existing motives' are, there will indeed turn out to be no external reasons. But that will only be because the set of my internal reasons has expanded to cover pretty well everything".[93] These readings of the internal reasons thesis present a problem for the integrity objection, as Chappell understands the objection as taking the form of an encounter with moral demands, which are not consistent with an agent's already existing motives. The narrow reading trivialises that encounter, as there is clearly nothing unduly demanding about the suggestion that agents might have reason to do what does not immediately

[90] Ibid.
[91] Ibid.
[92] Ibid.
[93] Ibid, 259.

occur to them. Whereas, the wide reading eliminates the encounter, subsuming moral demands into the motivations agents already have; if there is some essential human nature which calls out for impartial morality, instilling in each agent a motive to be moral, then moral reasons are not the products of external demands, but one's own projects.

3. Already-Existing Motives, Williams' View

3.1 Overview

Chappell has claimed that what counts as an already-existing motive is ambiguous, but that no matter how we interpret this ambiguity, no concept of what an already-existing desire is can capture what is problematic about the extreme demands of impartial morality. I challenge that claim on two fronts: that internal reasons thesis is not ambiguous in determining what counts as an already-existing motive, and also that Williams' integrity objection does not hang on the internal reasons thesis in the way that Chappell describes. This section will argue in favor of the first claim. In the course of making this claim, we will draw out a number of interesting features of the internal reasons thesis which Chappell neglects.

I will begin by addressing Chappell's two polar readings of what an already-existing motive is, the "narrow" reading, and the "wide" reading. I find that neither reading serves as a plausible fit for Williams' account of what the internal reasons thesis is, or what it was meant to accomplish. For the wide reading especially, the reasoning will turn out to be philosophically interesting, as it draws out the motivation for the internal reasons thesis: an emphasis on contingency in what agents can be rationally motivated to do. This contingency represents the animating motive behind many of Williams' philosophical contributions, and indeed is shared with the integrity objection.

Of course, the internal reasons thesis also aims to describe the relationship between an agent's motives and their reasons for action. And, it might be thought that if that relationship is *too* variable, that the internal reasons thesis might not be able to explain that relationship after all, nor could it justify defining limits on what *counts* as a reason or motive. The internal reasons thesis must then walk a thin line, ensuring that it identifies reasons which are responsive to an individual's personal character, both in that is genuinely reflective of the motives they actually have, and distinct from others who lack those motives.

This difficulty superficially resembles the worry presented by the "wide" reading Chappell offers, and so can be explored by examining what constitutes the *relation between* one's already-existing motives, and the array of reasons those motives give rise to. In the original formulation of the internal reasons thesis, Williams proposes two such relations as means of determining what an agent's reasons are, which are here called *instrumental reasoning*, and *persuasion*. In a later work, he clarifies that these relations are not distinct processes, but extremes on a continuum which characterises practical reason as conditioned by psychological facts, which complicates the way agents deliberate. All told, while the things which can serve as motives—and the reasons for action that follow from those motives—are rather diverse, Williams' commitment in the internal reasons thesis can be narrowed to the claim that the contents of an agent's pre-existing motivations determine whether or not they could be motivated to undertake some courses of action, if they are acting rationally. This is a notion he explored extensively in other works, which I use to explain what claims that the internal reasons thesis is attached to. In doing so, I demonstrate that the internal reasons thesis is not ambiguous in the way that Chappell suggests.

3.2 Two (and a half) Readings of the Internal Reasons Thesis

The point of offering both readings of the internal reasons thesis is to suggest that there is a lingering unclarity regarding what counts as an already-existing motive. Chappell is suggesting that Williams is not committed to ruling either of them out. What's more, if there had been enough evidence to collapse the thesis into one of these readings, then it would be clear that the internal reasons thesis cannot explain what is problematic about impartial morality. But in fact, neither of these readings is a viable interpretation of what Williams is up to when he introduces the internal reasons thesis.

The narrow reading Chappell offers is plainly inaccurate. That Williams believes that the motivations an agent has extends beyond the motives they are consciously aware of is proven by the fact that Williams thinks agents can be mistaken about what their true motives are. He writes: "*A* may not know some true internal reason statement about himself" because "*A* may be ignorant of some element of *S*"[94]. If an agent may be ignorant of one of their motives, how could it be that "at any time *t*, include little more than those motives that I am consciously aware of at *t*"[95]? He even more clearly rules this out later when he writes: "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."[96] Williams' claim is that an agent can discover their reasons and motives through deliberation, and so it must include some motives an agent is not immediately aware of.

So, an agent's internal reasons are not exhausted by the desires which occur to them at any given time. What of the wider reading? The suggestion here is that there may be a universal human nature which determines each agent's S such that all agents will end up with (at least

[94] Williams, "Internal and External Reasons", 103.
[95] Chappell, "Integrity and Demandingness", 258.
[96] Williams, "Internal Reasons and the Obscurity of Blame", 36.

some of) the same reasons. There are two ways that this wider reading can be taken as presenting a problem for Williams.

The first way, which I will call the strong-wide reading, suggests that human nature restricts agents' rational motivations to a restricted set of identical motives, so that the rational thing to do at any given time happens to be the same between agents. Williams would surely not admit of the strong-wide reading, as it defeats the purpose of arguing in favor of internal reasons in the first place. He writes: "internalist doctrine would be pointless if everyone's values, and everyone's *S*, were the same; in those circumstances, anyone's deliberation could be like anyone else's, and the distinction between externalism and internalism would fade away"[97]. If each person's motives happened to be identical, then each agent's motivational set wouldn't be *subjective* in the relevant sense.

The other way to understand Chappell's wide reading is to say amongst each agent's contingent motives, there is a set of universal motives shared by all agents. I will call this the weak-wide reading. According to the weak-wide reading, the contents of each agent's motivational set will resemble each other in including some universal moral commitments, which would vie with the agent's contingent motives to help shape the reasons that they have. For Chappell, this reading is problematic, because the internal reasons thesis is supposed to explain what is objectionably demanding in the integrity objection. And for Chappell, this demandingness derives from the externality of moral demands, that conceptualizes morality as demanding insofar as it asks us to act against our personal motives. But if morality is just one sort of motive among others, and so are not external to the motives which make up our

[97] Bernard Williams, "Values, Reasons, Persuasion" in P*hilosophy as a Humanistic Discipline*, ed. A. W. Moore, (Princeton University Press, 2006), 114.

motivational set, then we can listen and respond to moral demands without threatening our personal integrity.

If Chappell's worry is based on the notion that the internal reasons thesis assumes that all agents naturally have moral motives, then she is mistaken. The internal reasons thesis is about defining what reasons a rational agent has given the motives they already have, and Williams is convinced that an agent with only egoistic motivations are not necessarily irrational[98]. But whether or not all agents necessarily have moral motives is besides the point Williams hopes to make in the integrity objection. Williams's point is instead about the nature of those motives should they exist, it is about their relation to the other motives an agent has, and whether or not moral undertakings constitute sacrifice. Even if morality is an *acquired taste*, which agents only may become disposed to (much like other motives agents might develop under the internal reasons thesis), there is little in this thought alone to explain why moral motives are different in a way which explains their *demandingness*, as Chappell thinks Williams requires. And of course, if moral motives are motives like any other, it is hard to see how acting on moral reasons can be understood as interrupting an agent's ability to organize their own life.

This is a more plausible reflection of Williams' views, which might preserve the worry that Chappell presses on the integrity objection via the internal reasons thesis. On this account, the internal reasons thesis cannot explain what is unique about morality in that it is objectionably demanding if it cannot distinguish between morality and the other motives we do not typically think of as demanding. Without a strong distinction between moral motives and one's "already-existing motives", Chappell's sense of demandingness is lost.

It is true, however, that what might be called moral motives will conflict with other motives. And, assuming those motives somehow correspond to the directives of impartial

[98] Williams, "Internal and External Reasons", 112. See also "Egoism and Altruism".

morality, Williams has acknowledged an important difference between moral motives and other motives; moral motives sometimes purport to dominate other motives to the degree that morality sometimes asks agents to "step aside" from the rest of their lives in pursuit of some purely impartial directive. What's more, if morality is an 'acquired taste', it might be set against an agent's other, already-existing motives. This is not Chappell's reading of the integrity objection exactly, but the strategy is the same: if the reasons offered by moral motives are incompatible with an agent's already-existing motives, then agents are irrational to develop moral motives in the first place.

Chappell worries that if moral motives count as one sort of motive amongst others, "then the problem of demandingness will simply reappear as a problem about how to interrelate two different sorts of internal reason"[99]. Because she cannot ascribe demandingness to the case of an agent internally conflicted between two reasons which are genuinely motivating to them, she thinks that this means the internal reasons thesis fails to characterize the demands of impartial morality as demanding in the way that is required by the integrity objection. But since Williams thinks reasons for action only follow from motives an agent already has, Chappell focusing on the image of an agent deliberating over the purported reasons of impartial morality forces us to imagine an agent already somewhat motivated by impartiality. We don't need to say much about this agent, but given that they seem to be weighing their reasons, and some reason identified by impartial morality is among them, it appears more and more plausible that they have the requisite motivation. This is brought out even by the particulars of her strategy, where the consequences of impartial conduct are thought through, and found to be "pragmatically self-defeating". And so, the question becomes: given an agent's motivation to act impartially, are their supposed reasons to act impartially strong enough to overrule the reasons which follow from their other motives.

[99] Chappell, "Integrity and Demandingness", 259.

But perhaps that gets at the agent in too late a stage. Seeing that, for Williams, moral motives could be properly internal sources of reasons, we might learn less by examining the character of their instrumental reasoning given the motives they already have, and more by asking about the moment of *persuasion* which saddled them with the moral motive to begin with. Perhaps, the integrity objection can find its footing in a case of persuasion gone array, rather than in the pragmatic instrumental demands of moral projects.

3.3 Instrumental Reasoning and Persuasion

Under the weak-wide reading, moral motives *could* account for some of the motives an agent has, members of their subjective motivational set. But as is stipulated above, these motives will *not necessarily* be elements of any given agent's S. Along the lines of Chappell's objection, the weak-wide reading supposes that an agent might think they have reasons derived from impartial morality, but the reasons which follow from that motive undermine an agent's other, already-existing motives. This sets the explanation of the integrity objection away from questions about exclusively instrumental reasoning, and involves a conception of persuasion, which is how Williams describes an agent might deliberate their way into acquiring new motives. The tentative idea is that persuasion is a sort of deliberation, and so it can go awry, and saddle agents with merely *purported* motives, in the same way that Chappell's worry concerns purported reasons.

However, Williams does not admit of a sharp distinction between instrumental reasoning and persuasion, suggesting that purely instrumental reasoning always could result in the acquisition of new motivations, because of the psychological effects of deliberation. This section sets out to explain Williams' reasoning here, which will have the secondary effect of narrowing down the fundamental commitments of the internal reasons thesis. Subsequent sections deploy

this clarified version of the internal reasons thesis against Chappell's picture of the relation between the internal reasons thesis and the integrity objection, and ultimately show that Chappell's objection is misguided.

The internal reasons thesis commits Williams to the claim that a constraint on it being true that an agent has a reason for action is there being the appropriate member of S. However, not only are the contents of each agent's S highly contingent on their history and beliefs, but when an agent has reason to ϕ based on a desire D, "the ϕ-ing may be related to D only symbolically"[100]. As such, "the internalist account is generous with what counts as a sound deliberative route…[however if] we allow any extension whatsoever of the agent's S to count, we have lost hold of the notion of what the agent has reason to do in virtue of S"[101]. One might worry that if too few constraints are placed on what counts as sound deliberation, then each agent might be able to deliberate their way into just about anything. On the other hand, any constraint placed on sound deliberation is constitutive of a supposed rational commitment of the agent, a principle which informs the agent how they ought to see the relation between their actions and values. Such commitments would represent substantial assumptions built into the thesis, and if they were applied too liberally, these constraints might reduce the array of verdicts available in sound derivation, and in doing so eliminate the contingency and variation which motivated the internalist account in the first place.

Williams is prepared to admit of one such universal commitment among rational agents, that is to be "acting in the light of sound information… included in the *S* of every rational agent is a desire not to fail through error", in effect an agent's reasons must follow from a sound deliberative route from their existent motives.[102] Williams illustrates the effect of this

[100] Williams, "Internal and External Reasons", 103.
[101] Williams, "Values, Reasons, Persuasion", 115.
[102] Ibid, 111.

commitment with the image of a helpful interlocutor who assists an agent in determining what they have the most reason to do. The interlocutor is truthful and open about their procedure, and makes no attempt to lead the agent away from their original desires. On the contrary, the interlocutor will as best as they can make sense of the agent's S, making the satisfaction of the important contents of S through the agent's action their single priority, by helping the agent to ensure that they are acting on sound information in the pursuit of their desires.[103] As much as possible, the effect the interlocutor will have on the agent's deliberation should resemble the agent's ideal attempt at instrumental reasoning from their own S.

Should the agent reach a verdict with the interlocutor, and find that they have most reason to ϕ, must it be true that there will have been a sound deliberative route from the agent's already-existing desires to ϕ? Surprisingly, according to Williams the answer is 'no'! For example: "the agent might fall in love with [the interlocutor], something which radically modifies his *S*. He is led to new courses of action as a result of shared deliberation within the constraints, but the outcome cannot count as what he originally had reason to do."[104] Of course, the image of the interlocutor is only an illustration. However, the aim of the illustration is to show that even given minimal rational commitments assumed to be shared by agents, the act of deliberation might yet still vary following from contingent differences between agents. The example of the interlocutor demonstrates that the process which introduces reasons is not "algorithmic",[105] nor is it as Chappell supposes, "introduced only by conservative extension from the reasons that we already have".[106] Instead, there are "psychological effects of the agent's

[103] Ibid 116.
[104] Ibid.
[105] Ibid 117.
[106] Chappell, "Integrity and Demandingness", 258.

actual deliberating", such that "radically new material"[107] might manifest as the content of an agent's S.

Williams does not mean to introduce an exception to the internal reasons thesis. The new desires which may be introduced in the process of deliberation remain a product of the agent's original S[108]. For instance, the agent will not fall in love with the interlocutor for no reason at all, such a result will be the (albeit unexpected) consequence of circumstances meeting the agent's particular character. Though, that the additions to S are "radically new" is suggestive of an independence from the other members of S. We can understand this independence as these additions to S changing the course of that agent's deliberation, by sustaining reasons other elements of S would not sustain. To take an extreme example, consider what Williams called moral incapacities[109]. Moral incapacities represent limits of an agent's deliberation; an agent has a moral incapacity when it is not possible for them to form an intention to undertake some particular action, because of that agent's strong disposition to regard that action as immoral[110].

What is interesting about these incapacities is that they are not necessarily the straight-forward product of an agent's motives, rendered in a predictable decision following evaluation. Instead, moral incapacitates are the products of both *decision* and *discovery*. Decision, because one's moral incapacities follow from an agent's deliberation (like any reason for action), and discovery because such an incapacity might only emerge in contingent circumstances which combine various dispositions which the agent had not previously considered in conjunction with each other: "it presents itself to the agent also as a discovery, because the underlying dispositions have not before been focussed through and on to that very

[107] Williams, "Values, Reasons, Persuasion", 116.
[108] Ibid.
[109] Given the moral/ethical distinction, this might have been called "ethical incapacies", though I will not revise Williams' language here, for fear of muddling the issue.
[110] Williams, "Moral Incapacity" in *Making Sense of Humanity and Other Philosophical Papers, 1982 - 1993*, (Cambridge University Press, 1995), 51.

conjunction of features"[111]. An agent might find that they simply cannot form the intention to ϕ, even in the absence of any disposition regarding ϕ preceding deliberation. Indeed, it is the encounter with the circumstances surrounding ϕ which reveal its significance to the agent, a significance that might never emerge in an algorithmic, abstract assessment of what the agent has reason to do at *t*, given S.

This encounter is not with the undertaking of the action itself, but in the imagination of the undertaking in a determinate context. Just as in the image of the interlocutor, the act of deliberation which precedes ϕ has a palpable psychological effect on an agent, in this case introducing an element of S which can further ground that agent's reasons for action. This imaginative reasoning is Williams' *persuasion*, a way in which an agent might determine what their reasons and motives are, which is distinct from the instrumental reasoning which follows a sound deliberative route from desire to action. But, ultimately, Williams would not admit of a clear distinction here. The illustration of the interlocutor shows that even when restricted to a narrow procedure of purely instrumental reasoning, the agent engaged in reasoning is a human being, capable of determining and discovering—or discovering in the determination—novel motivational material which might radically change what they can be said to have reason to do. As Willaims concludes, instrumental reasoning "should not in fact be contrasted with persuasion"[112], as no deliberation from S can be purely instrumental without the possibility of the imaginative revision of the contents of S. It is clear that Williams does not think that there can be instrumental reasoning without imagination; he writes: "If someone is good at thinking about what to do, he or she needs not just knowledge and experience and intelligence, but imagination;

[111] Ibid, 52.
[112] Williams, "Values, Reasons, Persuasion", 117.

and it is impossible that it should be fully determinate what imagination might contribute to a deliberation"[113].

It is worth ensuring that this revelation about instrumental reasoning preserves the distinction between internal and external reasons. One worry we have been tracking has been that if there are too few restrictions placed on how an agent can satisfy a desire through action, then too many actions could be seen as following from from agents' S; agents' reasons would not seriously reflect the contents of their S, and so those reasons would not be distinctly *internal*. Of course, there is one at least restriction placed on practical deliberation: that the deliberation should be sound. But even with this minimal restriction, agents with similar motives might turn out to have very different internal reasons following deliberation. How can it be said that reasons are internal, when agents might unpredictably develop the prerequisite desire to an act they would otherwise have no reason to do?

But this would be to overstate the contingency Williams has introduced. The internal reasons thesis is reliable only if reasons are truly internal, meaning that the reasons that agents have are different from the reasons of others, on account of differences in their Ss. The internal reasons thesis is concerned with the necessary conditions under which we can admit an agent has a reason for action, and identifies cases in which some agent meets this necessary condition, and another does not. A strong example of such a case would be instances of moral incapacity to act in a way that others are manifestly capable of. Another example is that of Owen Wingrave in the thesis' original formulation. Owen just does not meet the necessary conditions to have a reason to join the army, though the rest of his family, were they in his position, would.[114]

[113] Williams, "Internal Reasons and the Obscurity of Blame", 38.
[114] Williams, "Internal and external reasons", 106.

To be sure, the point is not about the soundness of the internal reasons thesis, but about plausible interpretations of that thesis. Chappell claims that the integrity objection requires the internal reason thesis to explain what is so demanding about impartial morality, but that an ambiguity in the internal reasons thesis means that it cannot capture that result. As evidence, she offered some readings of that ambiguity, and argued that no matter how the ambiguity is read, the internal reasons thesis cannot explain why impartial morality is prohibitively demanding. In response, I first argued that neither of the interpretations Chappell offered are plausible readings of Williams. Then, I suggested that the purported ambiguity regarding pre-existing desires might be explained in the distinction between instrumental reasoning and persuasion. The 'weak-wide reading' could have been mobilized to attack the way an agent might, through persuasion, acquire moral motives corresponding to impartial morality. But here, I have argued that persuasion does not exist as a distinct process contrasted with instrumental reasoning, and so there is nothing in it to explain what is distinct about moral motives. So, we are left with a clarified view of the internal reasons thesis, but one which does not do the work Chappell thinks it must to support the integrity objection.

4. The Grounds of Reason

4.1 Overview

Chappell claims that the internal reasons thesis does not explain what is objectionably demanding about impartial morality in the integrity objection. I agree. However, I take issue with Chappell's reconstruction of Williams, and argue that notable ambiguities lead Chappell to the false impression that the integrity objection is about whether or not an agent has a reason to follow the precepts of impartial morality. This is wrong, not least because Williams affirms that moral systems might offer reasons for action to agents with suitable dispositions. Instead, the

integrity objection is not concerned with how an agent will arrive at reasons they think they have. I argue that the integrity objection is essentially about authority, about the assumption amongst some moral philosophers that the reasons offered by moral systems are uniquely decisive, especially over the other motives they have. In particular, it is an objection to the primacy of the moral over the ethical, of moral systems over the other commitments of agents.

4.2 Projects and Grounds

Chappell's reconstruction of the integrity objection begins with a view of what is involved in an agent having a reason to live, charting Williams' move from 1) the claim that the meaning in an agent's life, and their reasons to carry on living, depend on their individual ground projects and dispositions, to 2) the claim that it is absurd to demand that this agent acquiesces to the demands of impartial morality when this means abandoning these ground projects and dispositions[115]. But there is already an ambiguity in the reconstruction, that is whether the integrity objection relies on the agent's *ground projects* being under threat, or some other motive.

In Williams' parlance, *ground projects*, elsewise called *categorical desires*, are a subset of an agent's projects, distinguished by their unconditional significance from the perspective of the agent who has them. Williams contrasts ground projects with desires which are conditional on an agent being alive at the time of their satisfaction. One example of a conditional desire is the desire to satiate hunger: hunger is unpleasant, and one desire which motivates agents to eat is to avoid such unpleasantness. Typically, agents will be motivated to take on courses of action which guarantee their access to food. But if someone knew that they would be dead by Wednesday morning, they would not buy groceries for their dinner on Thursday night[116]. Not all desires are conditional in this way, however. He suggests:

[115] Chappell, "Integrity and Demandingness", 257.
[116] At least, not to avoid their own hunger.

> …it is possible to imagine a person rationally contemplating suicide, in the face of some predicted evil, and if he decides to go on in life, then he is propelled forward into it by some desire (however general or inchoate) which cannot operate conditionally on his being alive, since it settles the question of whether he is going to be alive. Such a desire we may call a categorical desire.[117]

Merrily, people often elect to carry on living, and there must be some motive behind that election. Williams accepts a liberal measure of contingency in the sorts of projects which can constitute such motives[118], but they are unified by the existential significance they have for the agent: in the pursuit of these projects the agent's relation to the world becomes a meaningful one, a sense that there is something in it for them in living.[119]

As Chappell reads Williams, she believes the integrity objection is pressing when an agent's ground projects are constantly under threat:

> The idea is that unless agents may have and be shaped by their own distinctive projects, "dispositions," and "commitments," they will have no reason to do anything…It is therefore pragmatically self-defeating for the extremist to try and impose moral demands that make it impossible for agents to have and be shaped by their own distinctive projects, dispositions, and commitments."[120]

However, neither of the cases in "a critique of utilitarianism" have agents commanded to commit to a lifetime of impartiality, nor obviously risk losing their connection to their ground projects or will to live. Instead, only one of their projects is under threat, and the risk is their connection to the project itself, not their whole life. In *George*, George's project appears under the guise of an ethical distaste for chemical biological warfare, and an unwillingness to aid and abet such things. Though, as Williams stipulates, his accepting the job will benefit other aspects of his life, including the personal welfare of his wife and children. There is even an ethical case (in fact, even a moral case!) to be made for his acceptance (which he might find appealing): that it will

---

[117] Williams, "Persons, Character and Morality", 11.
[118] Ibid, 12.
[119] Bernard Williams, "The Makropulos Case: Reflections on the Tedium of Immortality" in *Problems of the Self: Philosophical Papers 1956-1972*, (Cambridge University Press, 1973), 95.
[120] Chappell, "Integrity and Demandingness", 257-8.

likely forestall the research as compared to if he refuses.[121] What is at issue here is not George's moral identity *vs.* a lifetime of impartiality, or even whether there might be a sound deliberative route to George taking the job. What is at issue is whether or not George can regard his own commitment against such research as only "one satisfaction among others"[122]. Another agent might be able to view George's dispositions in that way, viewing his personal values as instrumental to the greater impartial value, rendering it—from the utilitarian perspective—insignificant. The same is true in *Jim*. Jim's case is one of personal conflict, between his part to play in the killing of an innocent prisoner, against the apparent moral weight attached to sparing the lives of the other nineteen. But there is no suggestion that his *ground projects* are at stake in this course, nor the conduct of his entire life, all that is at stake is some of their moral dispositions, a special sort of motive Willaims has called *commitments*.[123]

Typically, an agent will have more than one ground project[124], and so even if a categorical desire of both Jim and George hangs in the balance, Williams does not seem to think that they risk the sort of alienation and existential apathy required by Chappell's reading. Instead, Williams wonders whether utilitarians can expect Jim to merely stand aside, and count his commitment as only one project amongst many, humiliating it by contrast with the obvious moral significance of the lives of the nineteen. But this is the perspective utilitarians must occupy if they are to identify what are "*obviously* the right answers".[125] And this is a perspective an agent with personal commitments cannot occupy.

So, the integrity objection is not about a threat to an agent's ground projects, but to their ability to engage with one of their projects which involve their ethical commitments. It is

[121] Williams, "A Critique of Utilitarianism", 98.
[122] Ibid, 116.
[123] Ibid.
[124] Williams, "The Makropulos Case", 86.
[125] Williams, "A Critique of Utilitarianism", 99.

remarkable that this does not resemble the relation with the internal reasons thesis Chappell supposes. For Chappell, what might appear as a reason from impartial morality can be shown to be not a true reason, through the internal reasons thesis. Ultimately, it culminates in the claim that there is no reason to act on the reasons offered by impartial morality, because (1) they are external to the true motives of an agent's *S*, and (2) because acting on the reasons of impartial morality will alienate the agent from the true motives of *S*. It is the internal reasons thesis which shows that these purported reasons do not meet the necessary conditions Williams sets out for there to be a true reason, as if an agent deliberated soundly, they would see that acting on the reasons of impartial morality is incompatible with their true motives. So, agents who think they have reason to be impartial utility maximizers are mistaken.

I think the relation between the internal reasons thesis and the integrity objection is different than that. The trouble with Jim is not that he is under the misapprehension that he has reason to shoot the prisoner because acting on that reason risks conflict with some other project he has (indeed, Williams thinks Jim *does* have that reason[126]). The trouble is that he is in conflict between two reasons, and one of those reasons comes from a place with purports not only to offer a legitimate ethical consideration, but to exhaust those considerations. Jim has commitments, and one of those commitments might even be to the greatest happiness principle, but he also has other commitments which will inform his deliberation.

This view takes it for granted that the reason offered by an impartial moral system *could* be taken as internal reasons for action for an agent. Chappell does not accept this supposition, as she holds that the objection is about the encounter with an external source of norms. This is the basis for her view of the integrity objection, as a lifetime of impartial morality necessarily means a lifetime of an agent acting as though they do not have the personal dispositions they have. But

[126] Ibid, 117.

even utilitarian reasons can be internal reasons, and so given the internal reasons thesis, it would be impossible to alienate oneself from their own desires by acting on ostensibly external reasons.

The point about integrity is not that there is no reason to do what might be best when considered impartially. The point about integrity is that it is not *obvious* that we ought to consider these cases impartially, as though there were no other legitimate reasons for action. Agents' commitments give them their sense that there is something worthy in their conduct[127], it is their "moral relation to the world" which is lost when their circumstances are viewed only from the perspective of impartial prescription[128]. The point, later articulated against a different version of utilitarianism, is that the philosophy simply cannot treat these dispositions as preferences which are instrumental to the overall good; they are guides to an agent's practical life and conduct—sources of reasons for fully rational agents. Williams writes:

> [An agent's commitments] constitute a way of seeing the situation; and you cannot combine seeing the situation in that way, from the point of view of those dispositions, with seeing it in the [impartial] way, in which all that is important is maximum preference satisfaction, and the dispositions themselves are merely a means towards that.[129]

Jim may well have good reason to promote impartial satisfaction, but such a reason is not so authoritative over each and every agent to the degree that all of their other projects, especially their ethical projects, are irrelevant. Williams considers a utilitarian agent under conflict: "how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because [of] how the utilitarian sum comes out?"[130]. Indeed, for Williams it is the other way around; their utilitarianism is one project among others, but world satisfaction is not the currency from which their other commitments loan their value. Utilitarian philosophers misunderstand agency when

[127] Williams, "The Makropulos Case", 95.
[128] Williams, "A Critique of Utilitarianism", 104.
[129] Williams, "The Structure of Hare's Theory" in *Philosophy as a Humanistic Discipline*, ed. A. W. Moore, (Princeton University Press, 2006), 80.
[130] Williams, "A Critique of Utilitarianism", 16.

they fail to see the relation between an agent's commitments and their reasons, that an agent is more than a capacity to deliberate; they have projects, desires, dispositions and commitments; that they have integrity.

5. Conclusion

The mobilization of the internal reasons thesis to explain a feeling of alienation caused by impartial conduct is puzzling for a number of reasons, chiefly because a central lesson of the internal reasons thesis is that a rational agent can only follow reasons which are internal to them. So, how could the internal reason thesis explain the alienation felt by an agent who was motivated by the precepts of impartial morality? Chappell thinks that the integrity objection depends on an encounter with an external body of norms, but the truth is that Williams does not see impartial morality as a set of external reasons, chiefly because rational agents manifestly *can* act on reasons given from an impartial perspective. And, if a rational agent can act on a reason, then that reason is, in Williams' view, an internal reason. Nevertheless, Chappell mounts an interpretation of the integrity objection which depends on the internal reasons thesis, as it appears to her that there can be no sound deliberative route to impartial conduct from a set of desires that would be set aside by such conduct. But as she argues, the internal reasons thesis employs an all-too obscure notion of an "already-existing" motive, and so that thesis cannot secure the results that are required to explain the integrity objection. So, Chappell concludes that the integrity objection is suspect, because to require a result which the internal reasons thesis is too ambiguous to achieve.

But Chappell wildly overstates the ambiguity in the internal reasons thesis. Both the "narrow" and "wide" readings she offers are inconsistent with Williams' explicit commitments, and exploration of related ambiguities offers a robust account of internal reasons which stays

clear of Chappell's worries. More importantly, the integrity objection does not rely on the internal reasons thesis in the way Chappell thinks. This is because the internal reasons thesis is Williams' story about how an agent might have a reason for action, while the integrity objection is much less an encounter with an external normative other, the integrity objection is instead about the fraught inner life of an agent in conflict. In a slogan, the internal reasons thesis explains the route from an agent's motives to their reasons, while the integrity objection is a claim about moving from an agent's many reasons to one determinate action. "Integrity" is Williams' word for the relation between an agent and their reasons, a relation which is demeaned when a set of moral prescriptions purports to offer an exhaustive list of an agent's reasons without respect for their dispositions.

This reading of the integrity objection parts with Chappell's reading in a number of crucial ways. I defend the objection from Chappell's charge that it requires a certain clarity which Williams has failed to obtain; both because Williams has achieved the sought-after clarity, and because the integrity objection does not require it anyways. Instead of involving an account of the pragmatic consequences of acting for reasons which are not really your own, I read Williams as making a subtler point about the relation between one's reasons. The condition of an agent in conflict is one which can not necessarily be resolved through rationality, at least not without being faced with the necessity to, at times, choose between motives which are infungible. I think, ultimately, this is what he detests about the prospect of impartial morality: the search for a higher principle which comprehends all your cares, enough to determine the conditions in which their neglect is not only negligible, but that negative emotion connected to their neglect is rationally incomprehensible. The point is not that Jim should not shoot the prisoner (he probably should), the point is that it cannot be obvious to George if he has the sort

of commitment that are typical among agents. Impartial morality, in Williams' view, resents that fact, or worse, it cannot comprehend it.

Conclusion

I hope I have provided an interpretation of the integrity objection which makes plausible claims about moral theory, and the way it must relate to ethical practice. An agent's commitments are their ethical motives, which will offer them reasons that cannot be dismissed or excluded by other considerations. If an agent has a commitment to some object, practice, or relationship, then their commitment to an impartial moral theory cannot command them to disregard that commitment—its presence will offer the agent a legitimate reason for action. Indeed, commitments may conflict, putting the agent in an ethical dilemma, where they must neglect at least one commitment in favor of another. But this choice will not be easy or obvious if they are truly commitments, and it will not terminate in the conclusion that only one commitment was normative. Commitments are a special class of motives, and integrity is the agent's relation to their motives. When an impartial moral theory cannot grasp that an agent is in an ethical dilemma, due to a conflict of commitments, it presumes to ignore reasons the agent genuinely has. Indeed, its impartiality to the agent's contingent commitments frames deliberation as though anyone could be that agent in the position of choice, and supposes that they only have reasons that anyone else would have in their position. It denies that they are that agent with their own motives; it denies them their integrity.

Chapter One sought to make sense of Williams' line of argument in "A Critique of Utilitarianism", and isolate what was essential to the integrity worry. While much of that paper focuses on features specific to utilitarianism, such as the doctrine of negative responsibility, I argued that his reasons can be divided into two main points: 1) the command to deliberate impartially alienates agents' integrity, in failing to recognise other completing and legitimate reasons that agent has, and 2) utilitarianism requires respect for integrity to be independently

plausible. But Williams pursues 1) at various stages in his career, and issues this worry from integrity to a variety of moral theories. I concluded that the integrity objection is one that Williams returned to, developing his more and more considered view on the place of integrity.

Chapter Two strove to distinguish the objection from one form of the demandingness objection. Scheffler understands the integrity objection as rejecting the burdensome demands of morality, when it commands agents to act in ways which cause them to neglect their projects and desires. But Scheffler's interpretation could not explain important features of some of the cases Williams used to explain the objection, including and especially that integrity is attacked even when one's commitments and morality call for the same action. In fact, the objection is not about demandingness, it is about distortion. The distortion is that an agent's reasons have authority only if they follow from an impartial moral system, when in fact an agent's personal commitments supply legitimate reasons for action. This is brought out most clearly in the "Jim" case, and also the "Rescuer" case, one which is described in a different formulation of the integrity objection. Scheffler's interpretation, and his agent-centered prerogative, fail to attend to these features of the case of Williams described it, and so misplace the crucial role that the authority of reasons plays in the objection. As in fact, the objection is a rejection of a certain sort of authority, one which can be framed as the authority of the moral over the ethical, of moral systems over an agent and their commitments.

Chapter Three contended with Chappell's interpretation, which understood the objection as claiming that there is no reason to follow the reasons of impartial morality. She achieved this result by claiming that moral reasons are external to an agent's pre-existing desires, and that the internal reasons thesis explains why there is no sound deliberative route from agent's pre-existing desires to moral reasons. But Chappell also argued that the internal reasons thesis was inadequate

to explain that result, and so the integrity objection itself is suspect. But for Williams, the motives for morality are just as internal as any other commitment or desire, and so there is no ambiguous route to morality via other pre-existing desires. If an agent is impressed by the rational grounds of Kantianism, they may develop a commitment to the moral theory, and find themself subject to Kantianism's demands. This motive gives them the reasons for action (in Williams' view) which Chappell works so hard to question. But if there are reasons for action to act at the behest of impartial moral systems, Chappell cannot see how the integrity charge works to show that these reasons are problematic. But these reasons are not problematic in themselves. Morality is a commitment for Williams, but the integrity objection reckons with the fact that it is one commitment among many. The problem is that moral systems sometimes do not recognize this fact, and deny the legitimacy of all other commitments. Much less a problem about whether or not an agent has reasons, the integrity objection is about how an agent is commanded to deliberate given reasons in conflict.

The objection nevertheless relies on some serious assumptions. Commitments, as Williams describes them, are inescapable, in that one cannot view them though they are not pressing concerns on one's own conduct. Like a general coherently incorporating tactical maneuvers into a strategic scheme, we cannot rationally reject our commitments in an executive *all-things-considered* level of deliberation: "But strategic and tactical thoughts, unless the general is in a muddle, do not conflict, nor is there any conflict between the activities of thinking in the one style or the other"[131]. But there may be room to doubt, here. An agent who has a demanding commitment, such as a challenging obligation to a loved one, might grow to resent it, and wish they were different somehow, while never feeling as though their obligation wavers. It

[131] Williams, "The Structure of Hare's Theory", 80.

may be that people can get into a muddle, and have at least mixed feelings about their commitments while maintaining them. This is something that Williams pays little attention to.

The more serious assumption involves commitments' role in life satisfaction. For Willaims, being engaged with one's commitments plays an important role in one's overall happiness, and while this idea might seem plausible and intuitive, it is a strong claim for which there is vanishingly little argumentation. Indeed, if the importance of commitments is supposed to protect them from being discouraged by utilitarianism, Williams might have done more to explain why a commitment to utilitarianism itself could not be sufficient to protect the satisfaction of an agent, especially if that agent is also engaged in various projects and relationships which are not of an ethical character.

Granted these stipulations, the considerations which give rise to the integrity objection paint a picture of ethical deliberation fraught with conflict. An agent's motives are many, and various in their sources. Some of these motives, such as desires, can be quite contingent in the ways they can affect deliberation. Desires are acquired, dropped, ignored, supplanted or indulged, all in contingent ways. Commitments—in Williams' view—seem to come with rules, in that there are ways that a rational agent will not be able to deliberate if they have a commitment. An agent's ethical commitments structure their reasoning in ways mere desires do not, pressuring them stringently to contend with their own values and attitudes. On Scheffler's interpretation of the objection, this calls for agents to be released from the obligations of morality, making deliberation and action less burdensome. This isn't right. In Williams' view, ethical life is sometimes so burdensome that agents cannot possibly satisfy every commitment which commands them. It is to see that an agent's character is a source of motives private to themself,

and motives with the force and legitimacy to not be simply trampled by the rational authority of moral theory.

Works Cited

Chappell, Sophie Grace. "Integrity and Demandingness." *Ethical Theory and Moral Practice* 10, no. 3 (2007): 255–65.

Scheffler, Samuel. The Rejection of Consequentialism : A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions. Revised edition. Oxford: Clarendon Press, 1994.

Williams, Bernard. "A Critique of Utilitarianism." In *Utilitarianism : For and Against* by J. C. C. Smart and Bernard Williams, 75–150. Cambridge University Press, 1973.

Williams, Bernard. "Egoism and Altruism." In *Problems of the Self: Philosophical Papers 1956-1972*. United Kingdom: Cambridge University Press, 2011.

Williams, Bernard. "Ethical Consistency." In *Problems of the Self: Philosophical Papers 1956-1972*. United Kingdom: Cambridge University Press, 2011.

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

Williams, Bernard. "Internal and External Reasons." In *Moral Luck: Philosophical Papers, 1973-1980*, 101–13. Cambridge University Press, 1981.

Williams, Bernard. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity and Other Philosophical Papers, 1982 - 1993*, 35–45. Cambridge University Press, 1995.

Williams, Bernard. "The Makropulos Case: Reflections on the Tedium of Immortality." In *Problems of the Self: Philosophical Papers 1956-1972*. United Kingdom: Cambridge University Press, 2011.

Williams, Bernard. "Moral Incapacity." In *Making Sense of Humanity and Other Philosophical Papers, 1982 - 1993*, 46–55. Cambridge University Press, 1995.

Williams, Bernard. "Morality and the Emotions." In Problems of the Self. United Kingdom: Cambridge University Press, 2011.

Williams, Bernard. "Persons, Character and Morality." In *Moral Luck: Philosophical Papers, 1973-1980*, 1–19. Cambridge University Press, 1981.

Williams, Bernard. "The Primacy of Dispositions (1987)." In *Philosophy as a Humanistic Discipline* ed. By A. W. Moore. Princeton University Press, 2009.

Williams, Bernard. "The Structure of Hare's Theory (1988)." In *Philosophy as a Humanistic Discipline*, ed. By A. W. Moore. Princeton University Press, 2009.

Williams, Bernard. "Values, Reasons, and the Theory of Persuasion (1996)." In *Philosophy as a Humanistic Discipline* ed. By A. W. Moore. Princeton University Press, 2009.