City University of New York (CUNY)
CUNY Academic Works

Dissertations, Theses, and Capstone Projects | Graduate Center

6-2016

Actions, Reasons and Self-Expression: A Defense of Subjectivist-Internalism about Reasons

Carolyn P. Plunkett
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

 Part of the Ethics and Political Philosophy Commons

Recommended Citation
Plunkett, Carolyn P., "Actions, Reasons and Self-Expression: A Defense of Subjectivist-Internalism about Reasons" (2016). *CUNY Academic Works.*
https://academicworks.cuny.edu/gc_etds/1269

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.

ACTIONS, REASONS AND SELF-EXPRESSION:

A DEFENSE OF SUBJECTIVIST-INTERNALISM ABOUT REASONS

by

CAROLYN PLUNKETT

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2016

© 2016

CAROLYN PLUNKETT

All rights reserved.

Actions, Reasons and Self-Expression:
A Defense of Subjectivist-Internalism about Reasons

by

Carolyn Plunkett

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

_____	_____
Date	John Greenwood, Ph.D. Chair of Examining Committee
_____	_____
Date	Iakovos Vasiliou, Ph.D. Executive Officer

Supervisory Committee:

Jesse Prinz, Ph.D. (Supervisor)

S. Matthew Liao, Ph.D.

Jennifer Morton, Ph.D.

John Greenwood, Ph.D.

Iakovos Vasiliou, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Actions, Reasons and Self-Expression:
A Defense of Subjectivist-Internalism about Reasons

by

Carolyn Plunkett

Advisor: Jesse Prinz, Ph.D.

The central question of my dissertation is: what makes it the case that certain considerations are reasons for acting? This is a question about the truth-makers of claims about reasons, that is, what makes it the case that one has a reason to Φ rather than Ψ. There are two leading camps in the philosophical debate devoted to answering this question: subjectivism and objectivism. Subjectivist theories hold that one has a reason to do something when one has a non-truth evaluable favoring attitude towards that thing, e.g. desiring it. Objectivist theories insist that one's desires are irrelevant to establishing the existence of reasons; that some action or desire is morally good or valuable is equally and universally reason-providing, whatever else individual agents happen to desire. I argue that all reasons for action are subjective; that, conversely, there are no objective reasons. After rejecting objectivism and providing a general defense of subjectivist views, I defend a more nuanced subjectivist-internalist position called Expressive Reasons. Subjectivist-internalism is the view that reasons are not only rooted in agent's desires, but also that it must be possible for a reason to serve as the basis for an individual's action if it is to be a reason in the first place. Expressive Reasons is the particular view that R is a reason for A to ϕ when ϕ-ing is an expression of soundly deliberating A's self; and A, under ordinary conditions, would act on the basis of R. I argue that Expressive Reasons has unique philosophical advantages over competing views of reasons, and that it has compelling practical advantages in how it directs us to respond to different others.

ACKNOWLEDGMENTS

I owe the completion of this dissertation and any success to the people and institutions mentioned here, and then some.

Always the cheerleader that I needed, Jesse Prinz was a wellspring of support. His suggestions and clear, constructive criticism were extremely helpful in the development of this dissertation. I am indebted to Matthew Liao for taking on a CUNY student and advising me not only on this paper, but also various other projects in bioethics. Matthew has been incredibly generous with his time, attention, and guidance. John Greenwood, Iakovos Vasiliou, and Jennifer Morton provided very useful comments as I prepared for defense, and made the whole process very pleasant.

I would not have made it this far without Katie Tullmann. From the pro-seminar our first semester to dissertation writing retreats in my last year, Katie coached me through graduate school. Her philosophical acumen is matched only by her kindness. Laura Kane has provided much-needed comic relief and commiseration, not to mention cafeteria cookies.

Colleagues at the New York Society of Women in Philosophy showed me that philosophy is done best in groups, around a table. I am forever grateful to Gina Campelia and Rachel McKinney for introducing me to SWIPshop and to the women of NYSWIP for creating a forum for doing philosophy that is not intimidating, where one can present ideas and expect serious feedback without being made to feel small. Thank you to friends in the Emotions Reading Group and various Writing Accountability Groups for providing similar spaces.

Parts of this project were presented at the American Philosophical Association Central Division Meeting in March 2016 and in the Philosophy Department's Dissertation Seminar. I thank audiences and colleagues for their attention and feedback. The City University of New

York provided research support through the Enhanced Chancellor's Fellowship (2010-2015) and the Dissertation Year Fellowship (2015-2016). The Bioethics Program at the Icahn School of Medicine at Mount Sinai provided support through their Ethics Fellowship, plus an active intellectual community.

I owe many thanks to Arthur Caplan and colleagues in the Division of Medical Ethics at NYU Langone Medical Center for giving me the best incentive to finish this dissertation. I am thrilled and grateful to have the opportunity to do work that I love with such smart, supportive colleagues.

My parents' loving support has never wavered. I admire so much their commitment to education and to their family. Thank you for the freedom, confidence, and security to pursue my dreams and ideas in graduate school and beyond.

Finally, I thank Sean for generously giving me the time and space I needed to work on this project. Thank you for not asking too many questions about philosophy. I love our tiny home and our life together.

Motivation for this project came from a desire to vindicate individuals who have the do the seemingly wrong thing just to support their families and to survive. This dissertation is dedicated to them.

TABLE OF CONTENTS

1. Overview of the Debate **1**

1.1 The Conceptual Space
- 1.1.1 "Humean" Theories of Reasons
- 1.1.2 Varieties of Internalism
- 1.1.3 Subjectivism vs. Internalism
- 1.1.4 Clarifying Objectivist-Internalism
- 1.1.5 Hybrid Accounts of Reasons

1.2 Structure and Methodology

2. Objectivist Reasons & Objections **20**

2.1 Objectivist-Externalism
2.2 Objections to Objectivist-Externalism
- 2.2.1 Contra-Argument from Intuition
- 2.2.2 Contra-Analogy Between Epistemic and Practical Reasons
- 2.2.3 Contra-Avoid a Regress
- 2.2.4 Contra-Argument from Metaphysics

2.3 Onto Internalism
2.4 Objections to Objectivist-Internalism
2.5 Final Thoughts

3. Why Subjectivism? **52**

3.1 Disanalogy to Reasons for Belief
3.2 Argument from the Constitutive Aim of Action
3.3 Evidence of Reasons
3.4 Avoiding Presumptuousness
3.5 Avoiding Alienation
3.6 The Pillars of Subjectivism

4. Ideal Advisor Accounts of Reasons **79**

4.1 Ideal Advisor Accounts of Well-Being and Reasons
4.2 An Externalist Account
4.3 Objections in the Literature
- 4.3.1 Ideal Advisor as Nomologically Impossible
- 4.3.2 Ideal Advisor is Unrecognizable

4.4 Alien Recommendations
4.5 Motivating Subjectivist-Internalism

5. Subjectivist-Internalisms **103**

5.1 Generic Internalism
5.2 The Conditional Fallacy
5.3 Can Internalists Avoid the Conditional Fallacy?
5.3.1 Full vs. Practically Rational Selves
5.3.2 Action Descriptions and Explanations
5.3.3 A Somewhat Less Idealized Account
5.3.4 Do We Really Need to Worry about the Conditional Fallacy?
5.4 Reasons as "Normativized" Explanations
5.4.1 Reasons as Explanations
5.4.2 "Normativized Explanations"
5.4.3 Sound Deliberation
5.4.3.1 Improved Information
5.4.3.2 Imagination
5.5 Manne's Modifications
5.6 Limitations: The Need for a Stronger Defense of the Explanation Constraint

6. Expressive Reasons **137**

6.1 Self-Expression & Action: An Argument for the Explanation Constraint on Reasons
6.1.1 Self-Expression
6.1.1.1 Self-Expression Shows Cares
6.1.1.2 Self-Expression Shows *One's* Cares
6.1.1.3 Self-Expression Can be Overt or Non-Overt
6.1.1.4 Self-Expression Does Not Require an Audience
6.1.1.5 Self-Expression Can be Successful or Unsuccessful
6.1.2 Actions Self-Express
6.1.3 Actions, Reasons and Self-Expression
6.1.4 The Advantages of the Argument for S-I from Self-Expression
6.2 Expressive Reasons
6.2.1 Sound Deliberation
6.2.2 Ordinary Conditions
6.3 Some Objections and Replies
6.3.1 A+ is Not Idealized Enough
6.3.2 The Problem of Self-Knowledge
6.3.3 ER is Incoherent
6.3.4 Certain Selves Shouldn't be Expressed
6.4 Not Anything Goes
6.5 Conclusion

Bibliography **175**

LIST OF TABLES

Table 1: Taxonomy of Views about Reasons **5**

Chapter 1: Overview of the Debate

It's a common occurrence. Teddy and George are arguing about what restaurant to go to on Saturday night. In defense of his choice, Teddy says, "I just read a great review of the oysters at John Dory Oyster Bar in the *New York Times*; let's go there." George responds politely, "I don't like oysters. I know a great Peruvian place, let's do that." Teddy reminds George, "But I don't like Peruvian." The debate continues. Teddy and George are not arguing about the attributes of these restaurants; they agree that one has great oysters and another serves Peruvian food. They are arguing, rather, about which of these facts is a reason to go to either restaurant. What provides a reason for Teddy – liking oysters – differs from what provides a reason for George – namely, liking Peruvian food. In this case, it is clear that what reasons each has for going to a restaurant depend on the eater's preferences.

Are there reasons for acting, though, that do not depend on an agent's desires, preferences, or motivations? Let's think about Katie. Katie does not want to donate a portion of her income to feed a faraway family, but many moral philosophers insist that she has a reason to do so anyway. In doing so, they invoke an objectivist theory of reasons. Whatever her prior desires, preferences, or motivations, objectivists hold that Katie has a reason to donate her money. The intrinsic goodness of helping others provides a reason for Katie, and indeed all of us, to donate, regardless of whether she cares to do so. The moral properties, e.g. goodness, of

actions or the objects of our actions generate our reasons – not our contingent desires or preferences.

Subjectivists about reasons, on the other hand, defend the view that there is no reason for an agent to act that does not depend in some way on her prior desires, motivations, preferences, etc. Put positively: what counts as a reason for you to act depends on what you desire, care for, or are concerned about. Without such prior attitudes to ground true claims about what you have reasons to do, inert facts about menu items or the moral properties of helping others are not reasons. According to subjectivists, a consideration or proposition has the property of being a reason when one has a favoring attitude towards that thing. In this way, reasons are generated by one's desires, preferences, concerns, likings, etc. Most subjectivists are careful to say that not just *any* desires generate reasons, only the subclass of those that survive scrutiny.

An important note needs to be made at the get-go to make clear what's really at issue in this debate. The views I canvas pertain to the nature of *normative* reasons. Normative reasons are contrasted typically with explanatory reasons. The former justify actions that agents perform in accordance with them. The latter identify features of an agent's environment or psychology that explain, though do not necessarily justify, why she performed the action that she did.[1] The guiding question of my project is, then, "What makes it true that there is a reason for A to Φ?" instead of "Why did A Φ?" To be clear, my use of 'reason' throughout means 'normative reason.'

[1] The modern distinction between normative and explanatory reasons dates back to eighteenth-century philosopher Frances Hutcheson. He said: "When we ask the reason of an action, we sometimes mean, 'What truth shows a quality in the action, exciting the agent to do it?'…Sometimes for a reason of actions we show the truth expressing a quality, engaging our approbation…The former sort of reasons we will call *exciting*, and the latter *justifying.* Now we shall find that all *exciting reasons* presuppose *instincts* and *affections*; and the *justifying* presuppose a *moral sense"* (1728/1991, p. 308)

This is a question about the truth-makers of claims about reasons, that is, what makes it the case that one has a reason to Φ rather than **Ω**. In framing the question this way, I already assume normative realism about reasons. Normative realism is the view that propositions about what gives us reasons for action can be true or false independently of how things appear to us now. It says that facts exist about what individuals have reason to do, facts that may be substantially independent of, and more normatively compelling than, an agent's occurrent conception of her reasons. We can be wrong about what we have reason to do, even when acting in accordance with what we take to be reasons. Since all views canvassed in this dissertation assume normative realism, they are best understood as offering truth-makers for claims about what we have reason to do, or truth-makers for reasons-statements like "A has a reason to Φ." Assuming normative realism means that I will not address arguments for or against either expressivism about reasons or reasons nihilism.[2] It does not entail, however, a particular view about how to evaluate truth-assessable reasons-statements.

Despite deep disagreement over the truth-makers of reasons-statements, most normative realists also agree that reasons count in favor of acting. But this achievement is also quite minimal, since asserting that reasons "count in favor" of acting does not settle how or why certain considerations, but not others, count in favor of performing a particular action. The central question of this dissertation can be clarified further: what makes it the case that consideration R counts in favor of Φ-ing? As a question about normative reasons, this question asks at the same time, what makes it the case that consideration R justifies Φ-ing? Or, to use Julia Markovits (2014) wording, what makes the case that consideration R "throws its justificatory weight" behind Φ-ing?

[2] See Gibbard (1990) and Blackburn (2006) for expressivist views. For an analysis and comparison of (certain) normative realist views and expressivism see Dreier (2015). For a discussion of reasons nihilism, or the view that there are no normative reasons, see Dreier (2005).

This question has concerned philosophers for centuries, if not millenia; normative reasons are of central importance to philosophical ethics. The problem at stake is thought to be that if subjectivism is right – the view that what we have reason to do depends on our antecedent desires and attitudes – it seems to be very much in tension with moral absolutism and some forms of universal moral rationalism. Moral absolutism says that some actions are morally wrong for any agent no matter what motivations and desires they have. Universal moral rationalism says that if something is morally wrong, then every individual has a reason not to do it. The subjectivist must admit that the fact of some action's or object's objective wrongness or goodness does not *by itself* constitute a reason to avoid or pursue that thing; what reasons we have for acting, rather, depends on our desires and concerns. In so doing, the subjectivist supposedly must admit that agents can lack any reason to be moral, or lack any reason not to seriously harm another person. This implication makes many people uncomfortable. Savulescu (2009) implies that subjectivism leads to a reductio when he says, "[I]f we happened not to care about human beings, or persons, we would have no reasons…to care about them. If parents did not care about their children, then they would have no reasons to care about them." He concludes, "[A]nything goes…or at least anything could go depending on what we happened to care about" (2009, p. 225). This conclusion did not seem as absurd to Hume, though, who famously insisted, "'tis not contrary to reason to prefer the destruction of the whole world to the scratching of my little finger" (Hume 1740/1978, p. 416). I aim to stake out a position on reasons between "anything goes" and moral absolutism. I ultimately defend a subjectivist-internalist account of reasons called Expressive Reasons. I argue not only that reasons exist only relative to some set of concerns or desires, but also that reasons must be able to serve as the motive or basis

for action. Even if Expressive Reasons is right, as I argue, it does not commit us to "anything goes."

1.1 The Conceptual Space

I've already introduced some technical terms only the introduction to the introduction of this dissertation. One can be "subjectivist" or "objectivist" about reasons, "internalist" or "externalist." These words are used inconsistently at best and interchangeably at worst in contemporary philosophical literature. Sobel (2011) recently criticized Parfit (2011) for conflating subjectivism and internalism. Markovits takes “objectivism” and “externalism” to refer to the same view (2015, p. 3). To countenance confusion in my understanding of the debate over reasons and my usage of these terms, in this section, I will map the conceptual space and draw distinctions between these views. As far as I can tell, this is the first time that theories of reasons have been mapped in this way. The chart below and my explication of the views are meant as corrective for much of the literature on reasons as well as a teaching tool.

The chart below lays out four available positions on the nature of reasons and how those positions differ. I believe these four positions capture accurately an array of diverse theories and provide a useful way to conceptualize the debate.

Table 1: Taxonomy of Views about Reasons

	Subjectivist	**Objectivist**
Internalist	Reasons do, or would, motivate action Reasons generated by agents’ desires	Reasons do, or would, motivate action Reasons generated by normative facts or principles
Externalist	Reasons need not motivate Reasons generated by agents’ desires	Reasons need not motivate Reasons generated by normative facts or principles

I've used very brief definitions in the chart to characterize subjectivism, objectivism, internalism, and externalism. I will provide context and justification for my interpretations below. In brief, subjectivist views are united in asserting that reasons are generated by agent's desires, while objectivist views assert that reasons are generated by moral facts or values. As such, subjectivist views can also be referred to as "desire-based" while objectivist views as "value-based." Subjectivist and objectivist have very different answers to the question: what do I need to know, and where to do I need to look, to determine what I have reason to do? That is because they have very different answers to the question: what are the truth conditions on R's being a reason? Subjectivists advise us to look to the desires or "subjective motivational set" of a person. Among the truth conditions on something being a reason is whether it bears some relation to the desires of a particular person. Objectivists answer instead that we look to the structure of rationality, the demands of morality, or other non-subjective considerations, since the truth conditions on something being a reason is whether moral facts, principles, or values demand our acting in some way. Objectivists insist that reasons hold universally; they whatever is a reason is a reason for all agents regardless of their subjective concerns.

Already this distinction lands us in the so-called "problem of gridlock" (Chang 2014). Subjectivists and objectivists don't simply disagree about what are the truth conditions of what it takes to have the property of being a reason. They further disagree on the source of normativity of reasons. Subjectivists and objectivists are embroiled in a deep disagreement over the nature and grounds of normativity. I can only point this out and hope that bringing this disagreement to light yields a more fruitful discussion of normative reasons.

Both subjectivist and objectivist theories of reasons can break down further into internalist and externalist accounts of reasons. Internalist theories say that reasons would serve as

the basis or motive of actions of more or less idealized agents. In other words, agents would be motivated to act in accordance with their reasons if they were in improved conditions. As we will see, what those conditions are differs from theory to theory. Externalists deny this claim. They say instead that a consideration's ability to motivate actual, idealized, or rational agents is irrelevant; the truth conditions of reasons-statements do not include serving as the basis or motive for action.

Putting together these positions, we wind up with four families of theories of reasons: objectivist-externalist (O-E), objectivist-internalist (O-I), subjectivist-externalist (S-E), and subjectivist-internalist (S-I). The rest of this section provides a rationale for the terms used to characterize these positions and expands upon their meaning.

1.1.1 "Humean" Theories of Reasons

In demarcating the conceptual space I address in this dissertation, it's helpful to note explicitly what topics, concepts, or thesis I will not address. First among them are so-called "Humean" theories of reasons and motivation. One of the greatest sources of confusion in the normative reasons space is the conflation of "Humean" theories of reasons, internalist theories of reasons, and subjectivist theories of reasons. The growing varieties and uses of internalism exacerbate the confusion. I eschew altogether the use of "Humean" to describe theories of reasons.

The Humean Theory of Reasons (HTR) is characterized as asserting: "If there is a reason for someone to do something, then she must have some desire that would be served by her doing it" (Finlay & Schroeder 2015). Even though it there is a debate over whether Hume endorsed this claim, this theory of reasons is called Humean because it rooted in Humean sentimentalism, the idea that morality (and reasons) depends in some way on the passions. This characterization of

HTR, though, is consistent with both a subjectivist or internalist theory of reasons. A subjectivist endorses the claim that someone has a reason if it stands in a particular relation to an antecedent desire (whether or not that reason is capable of motivating action). On the other hand, an internalist endorses the claim that someone has a reason if it would motivate action under certain conditions. Internalism is consistent with either subjectivism or objectivism about reasons. Since HTR is ambiguous between internalism and subjectivism, I will not use it to characterize and categorize theories of reasons.

1.1.2 Varieties of Internalism

It is important to distinguish from the get-go what I'm calling the reason internalist thesis from various other types of meta-ethical internalisms. Unfortunately, the same term is used to characterize a multitude of meta-ethical positions. Internalism, as I am using the term, posits an existence claim about reasons: it puts forward necessary conditions of a proposition about normative reasons. Darwall terms this kind of internalism "existence internalism" (1983, p. 54). Existence internalism says that one of the necessary conditions of the truth of reasons-statements is that the reason would serve as the basis or motive for action (under specified conditions). This view is contrasted with reason externalism, the view that at least some reasons there are exist independently of its ability to motivate or explain action; it can be true of A that he has a reason to ϕ even though A has no motivation in his motivational set that would, either directly or by some extension through sound deliberation, lead him to ϕ. Arkonovich (2013) calls the this kind of internalism "reasons-motives existence internalism." This term makes clear that what's at issue are the truth conditions on the existence of reasons, and that internalism ties an agent's reasons to her motives. I will simply use "internalism" throughout the rest of the dissertation to refer to this view.

The distinction between the existence claims made by reason internalism and reason externalism was introduced in W.D. Falk's 1947 paper, "'Ought' and Motivation." It received much more attention, though, following Bernard Williams' 1979 paper, "Internal and External Reasons" which was reprinted in his 1981 book *Moral Luck*.[3]

My thesis defends existence internalism about reasons, so conceived, but I point out here three different varieties of judgment internalism in order to make clear what I am not concerned about. Judgment internalists, sometimes called motivational internalists, generally are concerned to show that motivation is intrinsic to or a necessary component of (1) recognizing reasons, (2) making moral judgments, or (3) accepting moral considerations. It is a view about the nature of normative thought and language and says that nothing can count as genuinely normative thought unless it is related in some way to motivation (Darwall 1983). Judgment internalists about *reasons* say that it is necessarily the case that if a rational individual recognizes that she has a reason to perform an action, she would be motivated to take that action. In a similar vein, judgment internalism says that moral judgments, e.g. the judgment that giving to the poor is right, are intrinsically motivating. We need not posit some contingent psychological fact in order to explain the motivational force of judgments. *Moral* motivational internalists say something similar: rational agents must be motivated by moral considerations that they accept, e.g. considerations about an object's value.[4] All of these views contrast with judgment externalism, which denies the existence of a necessary connection between recognizing reasons or making moral judgments and motivation. A standard worry is that judgment internalism cannot permit the existence of the amoralist, a person to claims to accept that she has a reason to do something,

[3] Citations of "Internal and External Reasons" reference the reprint, not the original, in keeping with typical citations of the work.

[4] My taxonomy of motivational internalism follows Darwall (1983), Garrard and McNaughton (1998) and Lenman (2011). I do not mean to imply that these terms are used consistently across all philosophers.

or genuinely judges some action to be morally obligatory, yet has no motivation to act accordingly. Judgment externalists can easily account for the amoralist.

It is not my project to defend judgment internalism. I'm concerned with how and why certain considerations or facts are reasons for acting in the first place – existence claims about reasons. The causal mechanism that instigates intention-formation and action upon seeing some fact as a reason is another matter entirely.

1.1.3 Subjectivism vs. Internalism

Reason-motives existence internalism ties an agent's reasons to her motives. Subjectivism, on the other hand, ties an agent's reasons to her desires. They both make existence claims, insofar as they posit necessary conditions on the existence conditions for reasons, and the truth conditions for reasons-statements. And at first pass, they appear to say the same thing. To think that, however, is wrong. In various works, David Sobel (2001a, 2001b, 2009, 2011) painstakingly shows that subjectivism and internalism are distinct theses, even though they have been – and are – conflated in the literature. I draw upon his work to clarify the distinction between subjectivism and internalism.

A subjective reason to ϕ is a consideration that counts in favor of ϕ-ing in virtue of the relation it shows ϕ-ing to stand in to our antecedent ends. Subjectivism is another term for what Jean Hampton (1998) previously called "identification internalism." Hampton says that identification internalism is the view that, "an agent has a reason to x if an only x-ing is connected…with an internal feature of the agent" (1998, p. 58). I prefer to use the term "subjective" for this kind of view to distinguish it from internalism. A subjectivist theory of reasons says that facts about the desires, goals, ends, or other internal feature bestow upon certain considerations the property of "being a reason" rather than facts about how the world is

or facts about the psychology of rational agents. A reason is "subjective" to the extent that it claims that what makes it true that something is intrinsically reason-providing for one is the existence of some sort of favoring attitude one has towards that thing. By "favoring attitude" I have in mind psychological states such as liking, desiring, caring for, preferring, cherishing, feeling good about, and similar attitudes. By convention, philosophers tend to use "desires" or "subjective motivational set" as the umbrella term for this whole class. I will keep with this convention. The existence of some sort of desire explains why the agent has a reason to obtain or pursue some thing. At the same time, subjectivists need not claim that what it is to be a reason is nothing over and above being such as to satisfy a certain sort of concern. That is, we don't have a reason to do whatever we happen to desire. The presence of a desire generates a reason and explains why we have one. But it is not sufficient to establish the truth of a reasons statement.

In sum, subjectivism is a thesis about the tie between desires, broadly construed, and reasons. No defensible subjectivist view says simply that we have reason to do whatever we happen to desire. Defensible subjectivist views place constraints on what desires generate reasons. Subjectivists views differ to the extent that they differ about what desires generate reasons, though all views endorse some kind of idealizing process. To remain a subjectivist view, this idealization needs to be procedural rather than substantive. A procedural idealization requires that one's desires be, for example, internally coherent, hold up to full information, or based on accurate forecasts of what the option would in fact be like. Ideal advisor accounts of reasons, for example, say that agents have reasons to perform actions that their fully informed selves would want their non-ideal selves to perform under the present circumstances. What I'll call generic internalism, on the other hand, says that agents have reasons to perform actions that they would want to perform upon soundly deliberating over what to do. Though distinct

subjectivist theses about reasons, both of these employ procedural idealizations – what I'll call an "idealization process" – to determine which desires actually generate reasons. Procedural idealization contrasts with substantive idealization, which would requires that one desires those things that are in fact desirable. Below I characterize as objectivist those views that build in substantive idealization to existence claims about reasons.

As noted by both Hampton (1998) and Sobel (esp. 2001a), subjectivist views break down into externalist and internalist accounts of reasons. For example, ideal advisor accounts of reasons are subjectivist-externalist while generic internalism is subjectivist-internalist. The difference between externalism and internalism is whether ability to motivate action is among the necessary conditions of being a reason. Reasons-motive existence internalism says that a reason would motivate, at least under idealized conditions (either substantive or procedural idealization). Williams had this is mind when he popularized the term internal reason: “If something can be a reason for acting, then it could be someone’s reason for acting on a particular occasion, and it would then figure in an explanation of that action” (1981, p. 107). As implied in this quotation, what it means for a reason to be capable of motivating an agent is that the reason would serve as the basis for acting. Another way of saying this is that the reason would genuinely explain why the agent acts in a way that is intelligible to the agent himself. Reasons-motive existence internalism is not a thesis about what mental state actually does motivating, but rather about the tie between reasons and their serving as the basis for action. Externalism is the thesis that some consideration can be a reason for an agent, even if it is not possible for that agent to act on the basis of the reason.

I have one more important note about subjectivism and internalism. In keeping with philosophical convention, I will use the term “desires” throughout to capture a very large class of

attitudes, pro-attitudes, and dispositions that includes things like preference for, concern for, care for, liking, wanting, feeling positive about, and the like. One's "ends" are the objects of those attitudes. Williams called this inclusive set of attitudes one's "subjective motivational set" or "S" for short. I will use subjective motivational set, desires, concerns, attitudes, and similar terms interchangeably throughout this dissertation. I hope it is not a distraction. In Chapter Six, I will introduce as a technical term "cares." Though there may be important differences between and among pro-attitudes, desires, and the like, I do not think they are relevant to defending subjectivism. All of those states can generate reasons according to subjectivism, and be used to explain why some consideration, but not another, is a reason for action.

Additionally, I will generally not use the term "motivations," even when discussing internalism, in order to avoid debates in philosophy of mind over what kinds of mental or psychological states are motivations. When discussing motivation, I will instead use terms such as the "basis for acting" "motive for acting" to remain agnostic on whether the belief that some consideration is a reason motivates action or the reason-generating desire motivates instead. In any case, I believe that equating reasons with the "basis for action" or "motive for action" is a better characterization of the internalist thesis.

1.1.4 Clarifying Objectivist-Internalism

Compared with the confusion surrounding the varieties of internalism and conflation of subjectivism and internalism, objectivism is relatively straightforward. An objectivist theory of reasons asserts that certain facts about things in the world - be they about values, existence of moral properties, or human psychology - establish the existence of reasons. We need not consider facts about *individuals* in the world, or their subjective desires, ends, or motivations to

determine whether some consideration is a reason for action. Objectivism sees normative reasons as universal requirements.

Understanding objectivism becomes slightly more complex when we consider its externalist and internalist derivatives. Objectivist-externalism (O-E) places the truth conditions on reasons-statements outside the individual, and does not claim that reasons must be able to serve as the motive for actions. It says, briefly, that values generate reasons, and so facts about what is valuable (or, in a similar vein, facts about what is right, good, or rational) are the truth-conditions on whether we have a reason to pursue a particular object or perform a particular action, whatever else we may want. Parfit (2011) provides the most influential defense of objectivist-externalism.

Objectivist-internalism (O-I) is perhaps more confusing to understand, especially considering the historical conflation between internalism, subjectivism, and Humean theories of reasons. It seems inconsistent if not paradoxical to assert that a reason can be objective and hold universally for all agents while also maintaining that the truth of reasons-statements depends on "internal" features of the agent. But there is no paradox in asserting both things. Recall that internalism says that if someone has a reason to ϕ, then it follows by necessity that she *would* be motivated to some degree, in circumstances of a particular kind, to ϕ. O-I theories of reasons say that the "particular kind" of circumstances in which agents will be motivated are those in which she is substantively idealized in accordance with the demands of rationality and/or morality. In other words, agents have reasons to do what they would be motivated to do if they were fully rational, but where the demands of rationality are substantive, in line with the demands of morality, rather than procedural. This kind of internalism has been called "tracking internalism" (Sobel, 2001a, 2001b), according to which necessarily, one cannot have a reason without being

motivated to act accordingly, not because motivations are part and parcel of normative reasons, but rather because the motivations of fully rational beings necessarily track the independent truths about normative reasons. Thomas Nagel (1970, 1986), Steven Darwall (1983), John McDowell (1995), Christine Korsgaard (1996), Michael Smith (1994), and Julia Markovits (2014) all defend O-I accounts of reasons. They are united in claiming that we all have reasons to do what we would be motivated to do under conditions of full rationality, and that these reasons hold universally, even if our non-ideal, often irrational, selves cannot act accordingly. What "full rationality" consists in varies by theorist and moral theory.

1.1.5 Hybrid Accounts of Reasons

The four families of views about reasons explicated in this section are mutually exclusive. One cannot hold that one and the same reason is both a subjective reason and an objective reason, for example, given that the assumptions about normativity on a subjectivist view of reasons deeply conflict with the assumptions about normativity that support an objectivist view. Some philosophers, however, defend a hybrid view of reasons.[5] Hybrid views allow that there exists more than one kind of reason. They offer more than one way of understanding the source of the normativity of normative reasons and grant that both objective and subjective accounts of normative reasons have some truth to them.

While I think that hybridism potentially combines the best of both worlds, I am going to assume for the sake of this dissertation that only one family of views can be right. This is largely motivated by a desire to consider the tenability each family of views independently before considering whether we may need two of them. Furthermore, even hybrid accounts must choose among objectivist and subjectivist theories which to import to a hybrid account. I'll do my best

[5] See Chang (2013) and Behrands (2015) for recent defenses of hybridism.

to defend the view that we only need to defend subjectivist-internalism, specifically Expressive Reasons, in order to get most of what we, or at least I, want from a theory of reasons.

1.2 Structure and Methodology

In support of "Expressive Reasons," the distinctive version of subjectivist-internalism that I defend, I provide objections to three of the four positions represented in the chart above. I will provide an objection to a crucial premise of nearly all objectivist theories and then argue that the leading subjectivist-*externalist* (S-E) account ultimately undermines its own subjectivist commitments. I argue that neither objectivists nor defenders of S-E have satisfying responses to my objections. By this process of elimination, I conclude that S-I theories of reasons provide the best account of reasons. I proceed to develop a positive argument for S-I and for a specific S-I theory, Expressive Reasons.

Chapter Two, "Objectivist Reasons and Objections" provides an overview of leading objectivist views, both externalist and internalist. I will focus on prominent contemporary objectivists: Derek Parfit, Thomas Scanlon, Thomas Nagel, and Christine Korsgaard. The first two of these, I show, fall into the objectivist-externalist (O-E) camp, while the latter two endorse versions of objectivist-internalism (O-I). I will be arguing not against the positions of particular philosophers, but rather against a general tendency to assert that normative properties or normative principles themselves constitute universal reasons, attributable equally to all persons. Those views are united in denying that antecedent desires, cares, or concerns are among the truth-making conditions of reasons. In addition to raising objections to particular objectivist views, I raise the concern that, on the whole, objectivist views beg the question in favor of objectivist theories of reasons. Objectivism about reasons seems necessary because it is thought

that objectivism is true for morality to exist at all. But that seeming so is not an argument for objectivism.

Chapter Three, "Why Subjectivism?" switches to offense. I review five arguments that have been adduced in support of subjectivism. None of the arguments I present are irrefutable; the best I can do is show that subjectivism is the *more* plausible alternative to objectivism about reasons. The five arguments are (1) the analogy to reasons for belief; (2) argument from the aims of action; (3) evidence of reasons; (4) avoiding presumptuousness; and (5) avoiding alienation. Though each argument has its limitations, considered together, the five arguments canvassed in this chapter present subjectivism as better accommodating individual difference than objectivism and less metaphysically dubious. Together, these features of subjectivism provide the two most compelling reasons to support subjectivism.

Recognizing the limitations of these arguments is an expression of humility. I will not argue in Chapter Three that there are a priori reasons to accept subjectivism rather than objectivism. On the contrary, I think this is an intractable debate. This may seem an odd admission, but I think it's the job of a philosopher to be humble about the limitations of her arguments and strength of her claims. I will argue that subjectivism is right, for reasons that will become clear in Chapter Three and beyond. Even the best defense of subjectivism, however, is unlikely to convince the staunch objectivist. This is not because the debate is merely verbal, but because the objectivist will persist in thinking that it is nonsensical to ask of universal moral norms, "by why is *that* a reason?" That is, they will not agree that that is a meaningful question, even in the face of philosophical disagreement about what normative facts are, which facts are normative, and wide diversity in moral norms cross-culturally. As I argue in Chapter Two, objectivists assume that universal normative facts exist, and those just *are* reasons. Chapter

Three presents five arguments for why we should reject that assumption.

Chapter Four, "Ideal Advisor Accounts of Reasons," delves into one of, if not *the* leading subjectivist views, the Ideal Advisor Account (IAA) of reasons. IAA says that A has a reason to Φ when a fully informed version of herself, A+, would want her non-ideal self, A, to want to Φ in her actual circumstances given A's desires, concerns, cares, and the like. A's desires ground what makes something a reason for her, but the process of becoming fully informed ensures that A's reasons are based on full information about the objects of her desires. IAA falls into the subjectivist-externalist camp. That is, it is a subjectivist account of reasons –reasons are grounded in agents' desires – but it does not make the further claim that reasons must serve as the motives for particular actions. In other words, an agent can have a reason or action even if it would not serve as the basis for her action under any circumstance. After raising several objections already found in the literature on IAA, I will raise a novel objection, arguing that the IAA leads to results that undermine the theory's subjectivist commitments, insofar as A+ may recommend that A have desires utterly alien to her current motivational set. IAA leaves open the possibility that A+ will recommend that A has a reason to act contrary to her deeply held convictions and beliefs - those elements of her S unique to and constitutive of her identity. As such, IAA permits cases where A is alienated from her reasons.

Chapter Five, "Subjectivist-Internalisms," presents various iterations of subjective-internalist theories of reasons. S-I says, briefly, that an agent has a reason to what she would be motivated to do if she were suitably idealized. Critics of S-I argue that S-I cannot avoid the so-called conditional fallacy. The conditional fallacy says, roughly, that an idealized agent may lack reasons that her non-idealized counterpart indeed has; it is, therefore, incorrect to tie the non-idealized agents' reasons to what her idealized counterpart would be motivated to do. This point

has led to a proliferation of S-I theories that purport to skirt the problem. I will argue in response that to take seriously the conditional fallacy is to implicitly assume externalism. I then provide a positive argument for the so-called explanation constraint on reasons, the crux of S-I.

Chapter Six, finally, defends my positive account of reasons, "Expressive Reason." In short, I propose that R is a reason for A to ϕ when ϕ-ing is an expression of soundly deliberating A's self; and A, under ordinary conditions, would act on the basis of R. The connection between expression and action secures the explanation constraint upon reasons. As such, this formulation of internalism respects the internalists' explanation constraint on reasons. In addition, I provide responses to objections to ER and similarly situated S-I accounts of reasons. I countenance the "anything goes" problem by sketching an alternative picture of the goals of normative reasons and their relationship to morality. It's not the case that "anything goes" even according to ER. The comments at the end of Chapter Six are admittedly speculative, and deserve much greater attention. My hope is that the strength of the arguments for ER provides a secure foundation for exploring the implications of the view in future work.

Chapter 2: Objectivist Reasons & Objections

Objectivism about reasons is the view that the truth-conditions of reasons-statements depend on what is valuable, right, or ratioanl, rather than on the desires, motivations, or concerns of ordinary people. It holds that there are certain normative considerations that are reasons for all agents – universal reasons – both to have certain desires and aims and to do whatever might achieve these aims. Objectivism has obvious appeal, and as such, a long tradition within philosophical ethics. Nearly every, if not every, major philosopher has had something to say about reasons, and many of them are objectivists. No dissertation chapter could do justice to even a minority of those philosophers. My own sympathies are with objectivism's opponents, though, and, in the space I have, I will attempt to resist those who claim that truths about what reasons there are can be known independently of information about the desires and motivations of those to whom those reasons attach.

I will focus on prominent contemporary objectivists: Derek Parfit, Thomas Scanlon, Thomas Nagel, and Christine Korsgaard. The first two of these, I show, fall into the objectivist-externalist (O-E) camp, while the latter two endorse versions of objectivist-internalism (O-I). As I outlined in the previous chapter, the difference between externalism and internalism when it comes to reasons is that externalist accounts divorce reasons and motivation. Internalism, on the other hand, says that it must be the case that reasons would motivate (rational) agents. Within objectivism, externalist and internalist views track another important distinction – mind-

independent vs. mind-dependent universals. Mind-independent universals are entities that exist independently of minds, that is, independently of whether beings like us inhabited the world or not. Mind-independent reasons exist whether or not there are people to discover or act upon them. Mind-dependent universals, on the contrary, exist only because there are beings with minds like ours that stand in a particular relationship to objects and things in the world and perceive them as such. Our minds and psychology are structured in such a way that all beings like us have the same experience, e.g. of color. In the domain of meta-ethics, mind-dependence says that normative reasons hold universally, not because they exist independently of humans, but because rational agents are constituted in such a way that they must see certain considerations as having the property of being a reason.

Even making these important distinctions, for the most part, I will be arguing not against the positions of particular philosophers but against a general tendency to assert that either normative properties themselves constitute universal reasons, attributable equally to all persons, or that necessary features of the rational agent generate reasons attributable equally to all persons. Those views are equally united in denying that antecedent desires, cares, or concerns are among the truth-making conditions of reasons. The position I oppose will be this form of objectivism – a position which can sometimes seem to be the only way for morality to proceed if it is to exist at all, but which I believe is neither defensible account of reasons nor a practically useful one.

2.1 Objectivist-Externalism

Derek Parfit is probably the staunchest defender of objectivism about reasons, and his 2011 book *On What Matters* serves as a 1000+ page defense of the view. Most of this defense

consists in knocking down subjectivist theories of reasons. For now I will focus only on his positive arguments, though, ignoring some of his most infamous negative arguments. I will point out affinities between Parfit's account and Scanlon's account of reasons. Though there are important differences, I will argue that both are subject to the same two criticisms. First, we can always ask of objectivists: why does some fact about the properties of an object constitute reasons? There is no non-question begging way to conclude that certain facts about objects constitute facts about reasons. Second, attempts to assuage the skeptic typically default to a subjectivist view of reasons.

I'll start with Parfit. He says:

> Object-given reasons are provided by the facts that make certain outcomes worth producing or preventing, or make certain things worth doing for their own sake. In most cases, these reason-giving facts also make these outcomes or acts good or bad for particular people, or impersonally good or bad (Parfit 2011, Vol. 1, p. 45).

There are at least two ways to construe how mind-independent "reason-giving" facts generate reasons – what I'll call a Moorean account and a buck-passing account. Consider the case of deciding which of two possible actions one has reason to perform. Let's stipulate that action A will produce more happiness than B. The fact that A will produce more happiness gives us a reason to A. Thus, you have a reason to perform A rather than B. There are two ways to construe why you have reason to perform A rather than B. On the one hand, the natural fact that A will produce more happiness makes that action a *good* action, and this further evaluative fact – that it is good – generates the reason to perform it (see Moore 1903/2004). On the other hand, according to buck-passing accounts, the natural fact that one action will produce more happiness *at once* provides the reason to perform it and renders the action good. This is Scanlon's position, called a "buck-passing account" since it passes the normative buck from attributions of goodness to other properties (e.g. producing happiness) in virtue of which some object, act, or occurrence

is reason-providing (Scanlon, 1998). According to buck-passing, the property of goodness is not the reason-making property, but rather some other property of the action. Parfit seems somewhat agnostic between these positions, though admits a preference for the buck-passing account (Parfit, 2001). Ultimately, though, he thinks that both Mooreans and buck-passers are committed to the same view: object-given reasons are provided by the facts about the properties of those objects, rather than facts about the desires or concerns of the persons to whom those reasons attach. First I'll review Parfit's arguments in favor of objectivism, and then return to Scanlon's buck-passing account.

Parfit's view is intuitionist, whereby it's simply obvious that some objects have reason-providing properties. This forms the basis of an important argument for objectivist reasons: the **argument from intuition**. He says, "When we are aware of facts that give us strong reasons to have some desire [to perform some action], our response to these reasons is seldom voluntary. It is seldom true that, if we had wanted not to have such desires, we could have chosen not to have them" (Parfit, 2011, Vol. 1, p. 47). We cannot choose to un-believe that our hotel room is on fire, much as we cannot choose to un-desire to leave the hotel room. The property of being on fire is one that gives us decisive reason to leave the room, whatever else we may want or care about. The evidence for that is that we cannot choose but to act in accordance with it – we simply must. This fact – the room is on fire – is one, Parfit thinks, we just do see as reason-providing. He adds, it is a serious mistake to deny that a room's being on fire provides a reason to exit it. Similarly, Parfit asks, "Who could possibly deny…that the nature of happiness gives us reasons to want to be happy?" (ibid., p. 57). Consider the fact that taking some medicine will give you many years of happy life. Parfit asks: "How could [Bernard Williams] believe that…this fact does not count in favor of your acting in this way?" (Parfit, 2011, Vol. 2, p. 434). Not only does Parfit think it

will be obvious when we have reasons to act, but also he thinks it is equally obvious when we do not; claims such as "I have a reason to waste my life" are "clearly false" (Parfit 2011, Vol. 1., p. 86). The inescapability of acting upon learning certain facts – the room's being on fire, the pill will prolong a happy life, a child is drowning – is evidence that those facts are reasons, whatever else we may desire or be concerned about.

A second argument relies on an **analogy between epistemic and practical reasons.** He asks his reader to think about the differences between two kinds of epistemic theories. According to the reason-based theory, "we ought to believe what the facts that are known to us give us decisive reasons to believe" (ibid., p. 92). An example is evidence-based medicine. Clinicians regularly consider facts about a medication's efficacy and a patient's particular pathology. When there is decisive evidence to believe that a particular medication is efficacious in treating the patient's pathology, the clinician has decisive reason to believe it will be efficacious in treating her patient. According to "the belief-based theory," "we ought to believe whatever, after considering the facts, we would in fact believe" (ibid.). This says: consider facts about the medication's efficacy and about the patient's pathology. Whatever you conclude is what, in fact, you ought to believe. Parfit admits the "belief-based theory" seems plausible, but claims that this is only because most of us also believe that the facts that remain after we've considered them give us decisive reasons to believe their content. The physician has reason to believe the prescribed medicine will be effective because the facts give her decisive reason, not only because she actually believes it. But "belief based theory" implies that we have no such decisive reasons; it only seems plausible when we add in the assumption that the facts that survive consideration are facts that give us decisive reasons to believe their content. We might lack decisive reason to

believe that the medicine is efficacious for this patient, even if we considered the information correctly, e.g. because the evidence of efficacy is based on fabricated data.

Parfit draws an analogy between reason-based epistemic theories and objective theories of practical reasons and between belief-based epistemic theories and subjective theories of practical reasons. Very crudely, objectivist theories of practical reasons say we ought to perform those acts that natural, moral, or evaluative facts give us decisive reasons to perform. Subjectivist theories say that we ought to want to perform those acts that, after the considering the facts, we would in fact want to perform. Subjectivist theories seem plausible, according to Parfit, only when we assume, as we did in the "belief-based theory" case that, upon careful consideration, we will want to perform those acts which coincide with what acts we have decisive objective reason to want to perform. If we think that the "reason-based theory" of epistemic reasons is superior to the "belief-based theory" – as Parfit thinks we must – then we should also favor the objectivist theories over subjectivist theories of practical reasons.

Parfit further argues that another advantage of objectivism about reasons is that it **avoids a regress**. He says,

> According to objectivists, we have instrumental reasons to want something to happen, or to act in some way, when this event or act would have effects that we have some reason to want. As that claim implies, every instrumental reason gets its normative force from some other reason. This other reason may itself be instrumental, getting its force from some third reason. But at the beginning of any such chain of reasons, there must be some fact that gives us a reason to want some possible event as an end, or for its own sake. Such reasons are provided by the intrinsic features that would make this possible event in some way good (ibid., p. 90-1).

On the other hand, subjectivists have no answer to the question, "but why that?" According to objectivists, the answer to that question bottoms out at a fact about objects of our actions or desires, such as their inherent goodness. Subjectivists have no such answer, and can always raise

the question again: "but why that desire?" The fact that objectivism avoids this regress counts in its favor.

Finally, Parfit raises an **argument from metaphysics**. He knows all too well that one of the most common gripes against objectivist accounts of reasons – and realist views in general – is that it is committed to the existence of "queer" facts (Mackie, 1977). Parfit resists this charge, and provides a realist account of the metaphysics of moral facts that, he argues, is not at all "queer." He supposes that establishing the existence of irreducible normative properties is sufficient to establish the truth of claims about reasons, given that he is already committed to admitting that certain normative properties provide reasons. Since my objection to objectivism does not rest on the intricacies of his argument from metaphysics, I will not go into too much detail. I will paint his metaphysical in broad strokes, though, since this is where Parfit's view and Scanlon's account most obviously diverge.

Parfit starts his argument with:

> There are other abstract entities and properties that are not mind-dependent, or created by us. Though novelists invent fictional characters, and composers create symphonies, mathematicians and scientists *discover* proofs and laws of nature. It is harder to explain what is involved in the existence of such entities and properties, and how we can know truths about them. Reason-involving normative properties are, I believe, of this independent kind (Parfit, 2011, Vol. 2., p. 475).

Parfit's argument goes on to make an analogy between numbers or other abstract entities discovered by mathematicians and scientists and those normative properties that generate reasons. Normative properties are not natural properties that exist in the spatio-temporal world as we know it, yet are not "queer" aspects of a non-spatio-temporal part of reality either. The existence of normative properties has no ontological implications and need not have any metaphysical content whatsoever. Parfit invokes two useful analogies to explain. First:

> If nothing had ever existed in any ontological sense, there would not have been any stars or atoms, nor would there have been space, or time, or God. But it would have been true that nothing ever existed. As we can also claim, there *would have been* the truth that nothing existed in an ontological sense. This truth would have *existed* in a different *non*-ontological sense. (ibid., p. 485)

And second,

> …This is like the way in which a series of symbols written on some page may be a valid proof of some theorem. Though these symbols exist in the spatio-temporal world, their property of being a valid proof is not, I shall claim, a further natural property of this world. This proof's validity is not itself a normative property. But this validity has the non-natural normative property of giving us a decisive reason to accept this proof. (ibid., p. 486)

Just as validity has the non-natural normative property of giving us decisive reason to accept a proof, so normative properties give us decisive reasons to act. Parfit's analogies make clear that he does not equate normative properties with natural properties. Rather, normative properties exist in a different, non-spatio-temporal sense, in some sense supervening on the natural. Claims about those properties, i.e. that they provide reasons to act, are true in a non-ontological sense, just as the truth that nothing existed nonetheless exists.

Since this metaphysical story about the attribution of mind-independent facts is where Parfit's and Scanlon's objectivist positions differ the most, I will move on to Scanlon's defense of normative properties and objective reasons. I then raise objections to both accounts.

Scanlon is also committed to the mind-independence of the kinds of properties of objects that generate reasons. That some action has the property of producing happiness generates a reason when it is true that this property is a reason-generating property of actions, objects, or occurrences. We must ask further: when is it true that some property is a reason-giving one? The answer, according to Scanlon, is that it is true when the fact that some property is reason-providing cannot be reasonably rejected. Scanlon says, "The content of the morality of right and wrong is determined, in my view, by considering what principles it would or would not be

reasonable to reject from various standpoints" (Scanlon, 2002, p. 519). The "principles" in question are claims about whether certain properties are reason-providing, e.g. the property of producing happiness, or easing pain. Reason-providing properties may be either evaluative or natural properties. For Scanlon, it is true that some property X is reason-providing if the claim that X is reason-providing survives a process of "reflective equilibrium" whereby the principle is considered from various standpoints and not rejected from any of them. The truth of the claim that some property is reason-providing, thus, rests on no facts about the antecedent desires or concerns of the people who consider it, but rather on the fact that the claim survived reflective equilibrium. In that way, Scanlon's is a squarely objectivist position, he simply provides a different story about why it is true that certain properties provide reasons than Parfit does. Reflective equilibrium is a method of arriving at mind-independent truths much like scientific theory construction.

2.2 Objections to Objectivist-Externalism

2.2.1 Contra-Argument from Intuition

"We simply must Φ" is used by Parfit as code for "We cannot deny that there is a reason to Φ." When the room is on fire, we cannot deny that there is a reason to leave. Note that Parfit allows that there being a reason to leave is compatible with there also being a reason to stay, say to save a child who is stuck in the building. The reason to leave is generated, at least in part, by the feeling (or anticipation of feeling) pain from the heat of fire. We accept that we simply must avoid the feeling of pain. Or take, another example, we simply must pursue happiness; we cannot deny there is a reason to do so. According to Parfit, the one of the best argument for this claim is

that, we cannot deny that we have a reason to want to be in states that we like, and that we have a reason to avoid being in states that we dislike.[6] Parfit's argument for this appeals to intuitions.

But Parfit's appeal to intuitions is methodologically weak. One may deny that she "simply must" have a reason to leave the hotel room on fire, just as one may deny that she "simply must" give away part of her income in order to contribute to ending world hunger. This response to Parfit is motivated by a deeper worry about the methodology employed in philosophical debate. Where clashes of intuition undergird two diametrically opposed philosophical positions (e.g. objectivism and subjectivism about reasons), appealing to those intuitions will never work to convince the other side. This has landed in the "problem of gridlock" as Chang (2014) puts it. I cannot deny the appeal of Parfit's simple argument from intuition. Yes, it seems like we should run as fast as we can out of the burning room; indeed that we have a reason to. But that does not establish that an objectivist theory of reasons is the best or only way to explain why I have that reason. Philosophers should expect more than "you simply must" as the proof for an interpretation. I may be, proverbially speaking, throwing out the baby with the bathwater; I will inevitably appeal to intuitions to motivate subjectivism. I can only register the methodological worry in doing so.

2.2.2 Contra-Analogy between Epistemic and Practical Reasons

Like belief-based theories of epistemic reasons, subjectivist theories of practical reasons seem plausible, according to Parfit, only when we assume that, upon careful consideration, we will want to perform those acts which coincide with what acts we have decisive objective reason to want to perform. He accuses subjectivists of assuming objectivism, and then building a

[6] I'll come back to this point below, in section 2.2.3.

procedure that all but ensures rational agents will arrive at the same reasons defended by objectivists.

Parfit's accusation may be partly right. After all, subjectivists and objectivists alike start with similar intuitions in most cases. "Of course you have a reason to leave the room that's on fire!" "Yes! Jump in and save the drowning child!" But the fact of having shared intuitions about what we have reason to do does not reveal whether we have them because our intuitions are truth-tracking, or we have them because our desires and beliefs about the world overlap, often substantially. Objectivist and subjectivist theories may yield extensionally equivalent sets of reasons, and their proponents will agree, in many cases, especially for similar agents, about which considerations are reasons. An appeal to intuitions will not settle the debate about why those considerations are reasons, however. Subjectivists will still deny that the explanation for why there is a reason in a particular case is because "being a reason" is a universal property of certain objects whatever else particular subjects want.

2.2.3 Contra-Avoid a Regress

Parfit argues that objectivism is preferable to subjectivism because it avoids a regress. Subjectivists have no answer to the question, "but why should we want to ϕ instead of Ω" except personal taste, a matter of contingent desires. On the other hand, objectivists can appeal to non-contingent facts about ϕ-ing that make it the case that we ought to ϕ.

Parfit defines subjective theories as those that claim that, "our reasons for acting are all provided by, or depend upon, certain facts about what would fulfill or achieve our present desires or aims…Since these are all facts about us, we can call these reasons subject-given" (Parfit 2011, Vol., 1, p. 57). If I happen to desire to count blades of grass, subjectivists' theories will say that I have a reason to count blades of grass, since that will achieve a present desire. His concern is that

subjectivists theories must admit that there are no reasons to have the particular desires that generate subject-given reasons, and, so, no reason to satisfy those desires. The difference between objective and subjective theories, then, according to Parfit, consists in whether or not there are reasons to satisfy our desires.

Oddly enough, Parfit endorses the claim that what our contingent likes and dislikes indeed generate reasons. This is because pain and pleasure generate reasons. We have reasons to be in states of pleasure and reasons to avoid states of pain. Parfit calls states of pleasure and pain "hedonic likings and dislikings" respectively. He says, "What we dislike is some sensation. What we want is not to be having a sensation that we dislike" (Parfit 2011, Vol. 1, p. 54). Our likings in turn generate reasons insofar as they, "make some of our conscious states good" (ibid., p. 55). Parfit admits that we have no reason to like or dislike things, but the fact of our liking and/or disliking constitutes an objective reason to desire what we like, and a desire to avoid what we dislike. Some of us just like pistachio ice cream better than chocolate, and get more pleasure from readings novels to non-fiction. If we didn't like those things, we wouldn't have reason to want them or to do them. The pleasure we derive – from whatever we happen to like – generates objective reasons to want those things. Objective theories, then, according to Parfit, avoid a regress. Where subjectivist theories have no answer to the question, "but why should you want *that*," Parfit can say, "because it gives you pleasure, and you have a reason to want to be in states that give you pleasure." But, crucially, Parfit says, there's nothing about the activity itself that generates a reason, but rather the nature of pleasure generates an objective reason to pursue whatever it is you happen to like.

This story about how likings generate reasons sounds awfully subjective. Parfit disagrees; he thinks that the distinction he makes between likings and desires avoids subjectivism. He

thinks that he can maintain the thesis that we have reasons to desire certain objects, actions, or phenomenological experiences, but no reasons to like or dislike what we happen to like or dislike. In other words, there's no reason for me to like pistachio ice cream better than chocolate, but liking pistachio gives me a reason to desire it. "We simply must" pursue the things we happen to like and avoid the things we happen to dislike. In other words, we have reason to desire those things that we happen to like, since the fact that something is pleasurable is an objective reason to desire those things that give us pleasure. Liking something is an objective, not a subjective, reason to desire the kinds of things that we like.

We should press Parfit on this distinction, as Sobel (2011) does. Likings are, arguably, a kind of desire, or at least, among the kinds of conative states that subjectivists maintain are reason-generating. Parfit readily admits that are no (normative) reasons to like or dislike certain experiences, such as the taste of pistachio ice cream, or the satisfaction of reading a great novel. These hedonic states just happen to us. Desires, for Parfit, are meta-hedonic states. He says,

> When we are in pain, what is bad is not our sensation but our conscious state of having a sensation that we dislike. It is similarly good to have sensations that we like. Such *hedonic likings* or *dislikings* cannot be rational or irrational, since we have no reasons to like or dislike these sensations. We also have *meta-hedonic desires* about our own and other people's pleasures and pains. Such desires and preferences *can* be rational or irrational, since we can have strong reasons to have them (Parfit 2011, Vol. 1, p. 3-4).

The strong reason to have certain desires is the fact that we have a reason to desire those things that we happen to like. But arguably, liking something is a kind of desiring it. Pressing Parfit on the same point, Sobel (2011) says, "Our question is not whether likings are different from some kinds of desires. Surely they are. Our question is whether likings are a sub-category of desire. Likings and meta-hedonic desires could be different, yet both sub-classes of desire" (57). Parfit points the following differences between likings and desires. First, we can fulfill desires but not

likings. We want to be having certain sensations that we happen to like. When we successfully have the sensation that we desire, we fulfill the desire but we do not thereby "fulfill" a liking. It persists. Second, likings and dislikings do not aim at the future. We only like or dislike *current* sensations, not future ones. Those current sensations, though, generate reasons to desire some state of affairs in the future. But even if we grant these differences, Sobel argues, we've only shown that meta-hedonic desires and likings are different. Parfit has not made the case that likings are not desires of some other sort, something like a "pro-attitude" or "positive-valence-making-characteristic" or "interest." All of these kinds of experiences, in addition to Parfit's meta-hedonic desires, according to subjectivists, generate *subjective* reasons to pursue certain actions or have certain desires.

Remember what is at issue here. Parfit thinks that objectivism avoids a regress. Objectivism can appeals to non-contingent fact to pursue one thing and not the other, including the fact that you have a reason to pursue what is pleasurable. But at least in some cases, what is pleasurable, according to Parfit, is what you just happen to like. That I happen to like scuba diving – the pleasure I derive from doing it – is its good-making property. But even Parfit admits that don't have a reason to prefer scuba-diving to snorkeling. I just do. Part of what it means to like scuba diving is to want to do it; I do not need a further reason to want to do it other than "I like it," which is, to my ear, a subjective reason.

Parfit argues in **avoiding a regress** that objectivism can provide a satisfying answer the question, "but why do I have a reason to do *that*" but at least in some cases, even the objectivists' position bottoms at a contingent liking that we just happen to have. Ultimately, I have a reason to scuba dive because I happen to really like it, which is to say, for no particular (normative) reason

at all. If that is a problem for subjectivists, as Parfit argues, it is at least as big of a problem for Parfit himself.

2.2.4 Contra-Argument from Metaphysics

I discussed above two attempts to establish that moral facts generate reasons: (1) Parfit's argument for moral realism and (2) Scanlon's argument from reflective equilibrium. Whatever story one provides, however, about how to establish the truth and mind-independence of moral facts, it is yet another step to justify the claim that particular mind-independent moral facts are reasons. I will argue against both views with the same argument, concluding that the fact that some object action has a certain normative, evaluative, or natural property does not establish that the fact is a reason, unless you already assume that objectivism about reasons is true. It may well be there are objective moral facts about what is right or what holds value, but that alone does not establish that those are reasons. My argument is informed by Falk's arguments (Falk, 1948) against Prichard's theory - and similar theories – of moral obligation. Falk's is the earliest use that I found of the open question argument (Moore 1903/2004) against objectivist-externalist accounts of reasons. I'll very briefly review Prichard's view of ought claims, and then Falk's arguments in response in order to motivate my own argument against Parfit's and Scanlon's objectivist-externalist views.[7]

Prichard's view attempts to explain why what we ought to do, that is, our moral obligations, provide reasons for us to act in accordance with obligations. That some act has the property of "ought-to-be-done-ness," according to Prichard is an objective fact – one whose truth does not depend on the antecedent feelings or subjective desires of individuals. To come to

[7] Burke (1983) argues that Falk's argument against Prichard rested upon a misunderstanding of Prichard's view. Regardless, I think it can be used to mount a similar argument against Parfit and Scanlon nonetheless. My exposition of the debate recounts Prichard's account of moral obligation; it was this to which Falk took umbrage.

recognize that an act is a duty, one need only concentrate on the relevant non-moral facts, facts about the state of affairs or situation one is in. For example, if one is wondering whether he ought to return a borrowed pen, think about the fact that one promised to return it. The natural fact that one made a promise to return the pen makes a claim on her to respond in a specific way, and that the claim made is stronger than any competing claim thrown up by other features of the situation, e.g. that it is one's only pen. Prichard says, "Suppose we come genuinely to doubt whether we ought, for example, to pay our debts … The only remedy lies in actually getting into a situation which occasions the obligation, or – if our imagination be strong enough – in imagining ourselves in that situation, and then letting our capacities of moral thinking do their work" (Prichard, 1912, p. 37). According to Prichard, "the sense of obligation to do, or of the rightness of, an action of a particular kind is absolutely underivative and immediate" (ibid., p. 27). One ought to return the pen. This argument is a precursor to Parfit's intuitionist take on reasons described above.

There are certain features that Prichard's view shares in common with Parfit's and Scanlon's. The fact that some action has the property of "ought to be done-ness" for Prichard or "being a reason" for Parfit and Scanlon in no way depends upon anyone's feelings, desires, etc. about that fact. Such moral facts can be established with complete independence from our feelings, desires, concerns, etc. For Parfit, those facts track truths about the existence of evaluative properties, for Scanlon, those facts track principles whose truth cannot be reasonably rejected, and for Prichard, those facts about what ought to be done pick out acts that are required by certain situations, an objective standard of morality (see Falk, 1948, p. 125). Falk says of this family of views:

> Views like these seem the modern descendants of traditional views of a more full- blooded but in essentials similar kind: that a person's subjection to moral law consists in his subjection to

> demands to do or to forebear made on him by a deity, or society, or a confused mixture of these; and that his actions would be right or wrong in proportion as they conformed to this standard. For, whatever their differences, there is this much in common between all such views. They presuppose, not unnaturally, that when someone "ought" or "has a duty" he is subject to some manner of demand, made on him with-out regard to his desires; and they imply that this demand issues essentially from outside the agent: that, whether made by a deity or society, or the "situation" (if this means anything) it has an objective existence of its own depending in no way on anything peculiar to the agent's psychological constitution (Falk, 1948, p. 125-6).

Objectivist theories of reasons insist that one has a reason to do whatever one ought to do. It is in virtue of an act's having the mind-independent, universal property of "ought-to-be-doneness" that we have reasons to perform certain acts.

In response to Prichard's argument, Falk argues that facts qua facts are not reasons. When you have completely enumerated all of the factual properties of an act (be they evaluative, moral, or natural), you still have not specified why it is worth doing. This table has the property of being round – so what? This painting has the property of being beautiful – what now? That action has the property of "ought-to-be-doneness" – who cares? On objectivist views like those canvassed, the perception that an act or a desire has certain properties gives me nothing more than reason for believing that an act or a desire has a certain property. But a reason for acting is not evidence for merely believing something about an act; such a reason is supposed to offer grounds for doing and not merely for believing.

Falk points out that objectivist-externalist theories require a twofold justification. First, one must be convinced that the act or desire in question has the property of "ought-to-be-doneness" or some other evaluative or natural property that itself constitutes a reason. But, even if the skeptic was convinced, she might claim, "I see that the act has property x, I don't see why that is a reason." The act has not been shown to have value for her. If in explaining a reason, a further fact is provided, that fact is as much in need of a reason for acting as is the fact that some

act has the property of ought-to-be-doneness or other natural properties (see Burke, 1983, Chapter Two for a similar argument). On Scanlon's view, one might say in response to the doubter, "but one could not reasonably reject the claim that we ought to do those things which have the property of producing more happiness than available alternatives." To which the skeptic will ask, "but why is *that* a reason?" and so on and so forth.[8]

Objectivists fail to provide a satisfying response to the skeptic who asks why facts and further facts constitute reasons on the objectivist view. Objectivists can only say that certain kinds of facts *must* generate reasons. But this is to obviously assume objectivism about reasons, since the skeptic need not doubt the truth of facts in order to doubt the existence of reasons. She is simply asking why they must be co-extensive. Objectivists like Parfit and Scanlon have no satisfying answer.

2.3 Onto Internalism

Parfit's and Scanlon's objectivism is also externalist, insofar as they make no claims to the effect that reasons would motivate agents to act accordingly. Parfit says, "We have [normative] reasons even if we would *not* be moved or motivated to act upon them" (2011, Vol. 1, p. 110), implying that reasons need not motivate or explain our actions despite our "having" them. Furthermore, "When we have [a normative reason] reason, and we act for that reason, it becomes our motivating or explanatory reason. Be we can have either kind of reason without having the other" (Parfit 2001, p. 1). Scanlon seems perhaps more agnostic, leaving open that recognizing that one has a reason might be sufficient for motivation, but landing short of

[8] A similar argument is found in Foot (1971) "Hypothetical Imperatives…". Peter Railton also raises a similar challenge to objectivism-externalism with the example of Hume's knave in Railton (1986b). These are all versions of the open question argument found in Moore (1903/2004). The open question was originally put forward to refute claims that equate the property of goodness with some non-moral property, in an attempt to refute ethical naturalism. Similarly structured arguments are used to refute claims that equate reasons with other moral and non-moral properties.

insisting that it must. He says, "the fact that the state I call seeing something as a reason has cognitive content does not debar it from serving as the motivation for an action" (Scanlon, 2002, p. 507). "Does not debar it" is a far cry from "entails that it must" serve as the motivation for an action, though.

In order to avoid the criticism I raised above to these objectivist-externalist views, namely, that we can always ask of objectivists, "but why is that a reason?" other objectivists have defended internalism about reasons, the thesis that, among the truth conditions of reasons-statements is that reasons would serve as motives for action in certain circumstances. This thesis answers a slightly different question than proponents of O-E. Objectivist-externalists are keen to establish that the source of reasons is other normative properties, e.g. that something is good. Objectivist-internalists, on the other hand, ask: how should someone like us, namely, a rational person, respond this to the world? The answer to that question yields an objective account of reasons, insofar as it will make substantive assumptions about human psychology and rationality. The fact that fully rational agents would respond in certain ways to the world establishes that we all have a reason to act accordingly. O-I identifies certain things about rational persons that require us to respond to the world in certain way. It says that we have reasons to perform particular actions because we ought to act as we would be motivated to do if we were fully rational. What "fully rational" means according to O-I theories of reasons will be explained below.

According to O-I, reasons must be dispositionally motivational – able to motivate in certain circumstances even if, at a particular moment, an individual is not so motivated. In this way, the existence of reasons is mind-dependent, dependent on the structure and psychology of the rational mind. This renders it less mysterious and dubious why certain facts, but not others,

are reasons. Reasons differ from facts in their unique ability to motivate individuals and figure in explanations of their actions. While Parfit and Scanlon have no satisfying answer to the skeptic's question, "well, why is *that* fact a reason?" objectivist-internalists like Thomas Nagel and Christine Korsgaard have a ready answer. The facts that constitute normative reasons are those that would motivate agents to act in accordance with them under a particular set of circumstances, namely, those of full rationality. If some consideration would fail to motivate a rational person, then it is not a reason.

Thomas Nagel fits squarely in the "objectivist" box, and advances an internalist conception of a reason. Nagel's internalism is motivated, I think, by an intuition he shares with Parfit – sometimes we cannot help but respond to reasons. But rather than go one to defend on externalist conception of reasons, whereby reasons need not entail motivation, even though they should, Nagel develops a complicated psychology whereby they *must*. He's project is avowedly Kantian in this regard. In this section, I'll explain Nagel's view and the arguments adduced in favor of it. I will also show parallels between Nagel's view and Korsgaard's account of reasons. Both Nagel and Korsgaard defend objectivist accounts of reasons, but also permit the existence of subjective reasons. For my purposes, I will focus only on arguments in favor of the former; they both claim that at least *some* reasons are objectivist, and that, given our psychology, for us, those reasons are intrinsically motivating. Parfit and Scanlon make no claims about psychological concepts or the structure of motivation; the objectivist's further defense of internalism thus requires a different type of argument which defends not only claims about the kinds of objects which generate reasons, but also the kinds of beings who respond to them.

My interpretation of Nagel's position is based on his 1970 book *The Possibility of Altruism* and his 1986 book *The View from Nowhere*. Nagel advances a version of constructivism

about moral facts, which informs his account of reasons. Constructivism is the view that insofar as there are normative truths, for example, truths about what we ought to do, they are in some sense determined by an idealized process of rational deliberation, choice, or agreement (Bagnoli, 2015). Nagel's procedural realism accepts that, "there are answers to moral questions *because* there are correct procedures for arriving at them" (Korsgaard, 1996, 36-7). This renders Nagel's view – like Korsgaard's – a mind-dependent account of objective moral facts. For Nagel, objects, actions, or desires that are valuable are intrinsically reason-providing, insofar as value generates reasons. Nagel's account of objective value says that value is created by creatures in the world, creatures like us human beings, when we view the world objectively, or as he says, impartially. Values are *real* and *objective*, and they generate normative reasons for us to act in certain ways, independently of what we happen to desire. Nonetheless, objective value does not exist independently of valuers. Objective values exist within a way of understanding the world – the way a very specific (and arguably impossibly distant) kind of valuer understands the world. In this way, Nagel's account of normative properties is more like Scanlon's, though Nagel does not utilize reflective equilibrium to defend the truth of claims about those normative properties, rather, his argument relies on claims about the psychology of valuers.

In this way, Nagel sets out to defend the existence not of objective reasons *per se* but a "system of objective reasons…a description of certain structural conditions to which reasons must conform, and an account of what in the nature of human beings the presence of that structure depends on" (Nagel, 1970, p. 99). He thus defends his interpretation of the conditions that must be satisfied in order to say of certain considerations or facts that they are objective reasons. Those conditions include:

- The consideration is one which would motivate a person to act when that person engages with the world impartially

- The consideration indicates of some act, event, or state of affairs that it is objectively valuable when viewed impartially. Those acts, events, or states of affairs are *non-relatively* desirable, and so intrinsically motivating.

Because motivation is internal to the concept of a reason – built into the conditions upon which some consideration is a reason – this account is an internalist one.

Nagel's defense of the existence of objective reasons, or of a system which admits of the existence objective reasons, hinges on the "impartial standpoint." Nagel admits (somewhat begrudgingly, when forced – see "Postscript" in Nagel 1970, pp. vii-viii) that there is a parallel system of subjective reasons – those considerations that motivate a person to act when that person engages with the world *partially* or *subjectively*. Though Nagel had hoped to show that all subjective reasons are rooted in further, objective reasons, it is widely accepted - even by Nagel himself - that he failed to do so. Nonetheless, Nagel has not abandoned the idea that a system of objective reasons persists alongside the subjective reasons, insofar as we engage with the world both impartially and partially, respectively. Similarly, subjectively valuable goals, acts, occurrences, and objects exist alongside objectively ones. You may value sailing, while I value skiing, but we both impartially value pleasurable activities. Thus, the very same activity has both subjective and objective value, since it can be viewed both partially and impartially.

The partial perspective does not admit of objective values or reasons. We have to shift to an impartial or objective standpoint to recognize objective value in things, happenings, or actions. Reasons are, in turn, generated by value insofar as those in the impartial standpoint will be motivated to act by the recognition of value in those objects. Objects that are objectively valuable are so, "in their own right and not through reducibility to some other kind of objective fact" (Nagel, 1986, p. 139). The impartial standpoint is another way of looking at the world, devoid of the particular aspects of our perspective that are unique and instead looking at the

world through a lens of shared humanity. Nagel's realism thus departs in nuanced ways from Parfit's. Parfit believes that normative properties exist in a real, yet non-ontological space. Nagel does not admit that value exists as a "new aspect of the external world," but is willing to admit that there are facts about what we should do. He later says those facts do not "extend radically beyond any capacity we might have to discover it" (ibid., p.139), but that neither is their existence dependent on our seeing them in some way. There are truths about what we should do, and they are true in light of the fact that it is what we would be motivated to do once in the impartial standpoint. But they remain true regardless of whether we enter the impartial standpoint or not. Reasons are what *would* motivate if we entered that standpoint, but we need not enter that standpoint in order to establish that those reasons exist.

Like Nagel, Korsgaard is considered a procedural realist, but she lays out a different account of the procedure that rational agents use to arrive at normative facts and reasons. Like Nagel, Korsgaard advances an internalist conception of reasons. Korsgaard argues that reasons are supposed to be action-guiding; they must connect with motivation to act. Reasons are not inert facts as Parfit and Scanlon would have it. Moreover, like Nagel, Korsgaard allows that some reasons, perhaps even most reasons, are subjective. We have them in virtue of our contingent desires and practical identities. What reasons professors have differs from what reasons students have in virtue of their particular identities. Yet, at the same time, Korsgaard argues, there are certain reasons that we all have in virtue of also having a *moral* identity, in virtue of being rational creatures. Korsgaard's distinction between practical identity and moral identity tracks loosely Nagel's distinction between the partial standpoint and the impartial standpoint. Of course we all have different reasons because we are different people, but we are also alike in one important sense: we share humanity. That shared humanity – a shared

perspective and identity – provides reasons to perform or refrain from performing certain actions, whatever else we may want as professors, students, etc.

Importantly, though, Korsgaard does not think that objective reasons are responsive to value in objects, but rather the nature and value of humanity itself. In fact, Korsgaard is often considered a subjectivist about value (see Parfit 2011, Vol. 1., esp. pages 94-95), as holding the view that we confer value on objects by preferring them, rather than prefer them because they are objectively valuable. Of course, sometimes, and perhaps even often, we confer value on objects, but, as I've shown Nagel is also committed to the idea certain objects are valuable regardless of our liking or desiring them when we would be motivated to perform acts or pursue those objects from within the impartial standpoint. For Korsgaard, only humanity is intrinsically valuable, but that fact alone constitutes a reason to treat humanity – in ourselves and other "citizens in the Kingdom of Ends" - in certain ways.[9]

Korsgaard clearly states, "'Reason' means reflective success" (1996, p. 97). Reflective success is measured on the Korsgaardian picture by whether or not an action is consistent with our moral identity, that is, whether it is one that we would endorse not only as persons with practical identities, but also as "citizens of the kingdom of ends." An action achieves reflective success when it does not treat others as a mere means, but rather an end in itself, in accordance with Kant's Formula of Humanity, which Korsgaard takes to be the bedrock of a Kantian moral

[9] I interpret Korsgaard as saying humanity is intrinsically valuable, as when she says, "All value depends on the value of humanity; other forms of practical identity matter in part because humanity requires them" (1996, p. 122). One might interpret Korsgaard differently. She says that by "reasoning backwards" (Pauer-Struder & Korsgaard, 2002, p. 59) from any particular conception of our selves (e.g. as professors), we will arrive at the judgment that we value humanity. A conception of ourselves as valuable qua-human is inescapable. Wherever we start from, we will see that the value of humanity the source of all normative force, and that the value of humanity is nonrelativistic. Now, since the way of arguing that conclusion is "backwards" one might be inclined to say that Korsgaard places only extrinsic value on humanity. It is extrinsically valuable insofar as it has a particular function in our lives, namely, the function of conferring value on all other things (See especially Korsgaard 1996, p. 122-3). Either way, we end up with the conclusion that that value of humanity is non-relativistic, the source of all other value, and necessary.

system. The Formula of Humanity is a command of reason, one that cannot be denied by fully rational agents. We cannot be rational and yet not motivated to act in accordance with the Formula of Humanity. Thus, we have a reason to do so. The nature of humanity gives us a reason to follow this principle and perform actions that are in accordance with the Formula of Humanity. The nature of humanity is such that we are capable of endorsing our reasons to act, a necessary condition for performing intentional actions. That capacity renders humanity at once the source of normativity – we give ourselves our reasons upon achieving reflective success – and itself places objective constraints on what considerations will achieve reflective success as reasons. Only those considerations that accord with the Formula of Humanity will achieve reflective success, and so only those considerations are (objective) reasons (see especially Pauer-Struder & Korsgaard, 2002).

Korsgaard also thinks that our reflective endorsement of reasons for action is intrinsically motivating; we are motivated by our awareness that our action is good, one which coheres with demands of humanity. As beings with a certain moral identity, we will be motivated to act in accordance with normative reasons. To be a reason, according to Korsgaard, is just is to be a normative force, one that motivates persons in virtue of their recognizing it as such. Korsgaard argues this must be so in order to make sense of the idea that reasons motivate and explain actions, in addition to justifying them (see especially Korsgaard, 2008, pp. 208-9). They do so because we are already rational, self-conscious individuals. Responding to reasons "out there" does not render us rational; rather, we respond to reasons because we are, fundamentally, rational beings. Our own nature and humanity both generates reasons to act and itself constitutes a reason to perform – and refrain from performing – certain actions.

Korsgaard, like Nagel, thinks we have reasons to treat humanity in certain ways regardless of whether we in fact recognize ourselves as "citizens of the kingdom of ends." The relevant fact is that we would be motivated to act in certain ways (e.g., in accordance with the Formula of Humanity) if we were fully rational. The method by which we discover reasons is similar on both Korsgaard's and Nagel's theories. Nagel asked that we zoom out of our partial perspectives and consider the world from an impartial perspective; what would still motivate us from that perspective constitutes an objective reason. Korsgaard asks that we reason from our practical identities – as professor, student, friend, and spouse – to the conclusion that humanity is a necessary condition for doing anything, the source of our reasons and obligations. We have reason to act in accordance with those considerations that would motivate us from the standpoint of our moral identities. To act from our practical identities – to act as a professor – without deliberation as a moral agent – is "evil" according to Korsgaard. She expresses this sentiment quite clearly here: "If an agent consciously and reflectively decided to treat a contingent practical identity as giving him a reason that is ungrounded in moral or human identity…then he would be evil" (Korsgaard, 1996, p. 250). When we deliberate over what to do, we must do so not only as professor, student, friend, or spouse, but also as "citizens of the kingdom of ends" with a certain moral identity independent of those practical identities. We therefore must transcend the standpoint of our contingent concerns as beings with practical identities and consider what considerations would be motivating for perfectly rational, moral agents instead.

According to both Korsgaard and Nagel, finding what one has reason to do is a matter of following a particular procedure to arrive at the objective, impartial, and universal moral point of view. That moral point of view is reached by abstracting from a personal point of view.

2.4 Objections to Objectivist-Internalism

The similarities between Nagel and Korsgaard render them vulnerable to the same following two objections. The first, which I call the **view from here** objection, argues that the morality cannot be adequately understood as a matter of abstracting from personal points of view. The juxtaposition of the moral point of view and the personal view does not capture the way in which our personal lives are imbued with moral considerations. We need not – and, indeed, cannot – transcend the personal view in order to be morally motivated. Susan Wolf raised this objection in "Morality and the View from Here" (1999). The second objection, called **the subjectivity dilemma**, argues that Nagel and Korsgaard's positions present the following dilemma: Either, like Parfit and Scanlon, objectivist-internalists beg the question in favor of objectivism about reasons, or they advance a subjectivist theories of reasons. Followers must decide which horn of the dilemma they would prefer to embrace.

In response to views of morality generally and normative reasons specifically that hold that objective values and reasons are those values and reasons that would motivate us from a moral point of view, Susan Wolf argues that this view of morality "arises from and perpetuates a false picture of human psychology and value, and it encourages an unduly narrow and ultimately implausible conception of what a correct and rational morality might be" (Wolf, 1999, p. 204). Morality – and what we have reason to do – is rooted in the **view from here** rather than the view from nowhere, as Nagel and Korsgaard would have it. Wolf thinks the intuitive support for the moral point of view accounts of reasons comes from too narrow an understanding of what a personal point of view encompasses. A personal point of view – the view from within our practical identities and partial perspectives – is typically described as inherently egoistic, concerned only with satisfying selfish desires and interests. It is not surprising that Nagel held

that altruistic behavior is impossible from within a personal perspective so-construed. The moral point of view captures the demands on us from outside, insofar as from that point of view we would be motivated by the value of objects and other persons and rational principles, rather than motivated by our subjective and personal desires. A view that distinguishes between a moral and personal point of view maintains that we are incapable of recognizing the value of other objects and other persons, or acting in accordance with rational principles, from within the personal point of view. Yet, Wolf argues, that assumption is false. She asks us to consider the mother who stays up all night in order to sew her child's Halloween costume. She is responding to the needs of others in way that sacrifices her own interests, but her motivation and justification for doing so do not rely on a moral point of view. She does not do so because it is objectively best or morally required; she does so because of an attachment to her son. Insofar as a clear distinction between moral and personal points of view is untenable, theories of reasons that say that what reasons we have from within those points of view are distinct kinds of reasons are also untenable.

Nagel and Korsgaard, and other objectivist-internalists might respond here that what they really advocate for is a hybrid view. Insofar as both allow the existence of subjective reasons, and that we are often motivated by elements within a personal perspective, they really need only say that in some cases we have both objective *and* subjective reasons to perform some action, e.g. helping a friend. The fact that our relationship to a friend does most of the motivating does not vitiate the existence of an objective reason to do so, too. Helping friends really is objectively valuable; it really is a way of respecting the humanity of others. As such, there is, in addition to a subjective reason to help, an objective reason to do so as well. But Wolf thinks even this combination view is false. What actually motivates us is not a detached, impartial consideration

of the value of objects and those around us, but our particular relationships with others and objects. When we are considering whether or not to help a friend, deliberating from the moral point of view does nothing to change or enhance the motivation we already have to help in virtue of the particular relationship we have to our friend. That our action is consistent with treating her as an end in herself is irrelevant to whether or not we are motivated to help her and irrelevant to the justification of such an action. That helping friends is also objectively valuable from an impartial perspective – that is, regardless of who the friends happen to be – lends no extra oomph to your motivation or justification for helping *your* friend. The hybrid view leaves out the specificity of our attachments to the things we actually value. Wolf says, "Though the things we care about may be both personally satisfying and morally valuable, we are engaged with them not under these general descriptions or in proportion to their general value of either sort" (1999, p. 212).

This point leads me to the **subjectivity dilemma** objection to objectivism-internalism about reasons. This objection points out that there is no stable objectivist-internalist position. Those who fall into the O-I camp should either embrace externalism about reasons, or embrace subjectivism. Given the problems with objectivism-externalism canvassed above, they should do the latter.

If, as Wolf says, none of the motivational force in helping a friend is coming from considering the case from a moral point of view, there is, then, no objective reason to help her on an internalist conception of reasons. The fact that some action accords with the Formula of Humanity or is valuable from an impartial perspective is not doing the motivating in these cases. An internalist conception of reasons holds that among the truth conditions of reasons statements is that the reason would motivate. Reasons must motivate the fully rational person. So, if some

consideration does not motivate rational agents, then it is not a reason. In cases like helping a friend, considerations about the objective value of helping is doing no motivating, despite it being entirely rational to do so. The supposed fact of its objective value is, then, not an internal reason. Even if certain facts remain – that some action would motivate me if I were in the impartial standpoint or that it is consistent with the Formula of Humanity – those facts are not necessarily reasons. The person who helps her friend can still ask, "but why is the fact I would be motivated to help my friend if I were reasoning like that a reason?"

If reasons must motivate, but what does motivate in most cases is a matter of our personal relationships and subjective values, then facts about what would be valuable from a moral point of view or would be rational from a moral point of view are just inert facts. The same objection that Falk raised against Prichard, and I raised against Parfit and Scanlon, holds against objectivist-internalist views, too. Nagel and Korsgaard must assume objectivism about reasons in order to for it to be true that facts about value and rationality are reasons. In advancing an objectivist-internalist view, they must assume that certain facts would motivate the fully rational person, even if, in practice, they are not. We can ask of facts about what would motivate us – so what? Why is that a reason? Unless you assume objectivism about reasons, those sorts of facts are not necessarily reasons. So, the first horn of the subjectivity dilemma is that objectivist-internalists like Nagel and Korsgaard simply become objectivist-externalists, insofar as they are forced to admit that even from the moral point of view, what actually motivates us typically has more to do with personal relationships and desires than the impartial value of humanity.

Alternatively, Korsgaard and Nagel and other objectivist-internalists could embrace a different horn of the subjectivity dilemma. They could say that the moral point of view is not a transcendent point of view, but rather just another point of view. We are professors, students,

sisters, friends, and moral agents. We can deliberate about what reasons we have from any of those perspectives, and what reasons we have will vary in accordance with what's important when occupying any one of those roles. Aside from the oddness of juxtaposing a moral point of view with that of being, say, a parent, insinuating that one cannot deliberate morally about what do qua parent, the idea that the moral point of view is like any other we may adopt renders the reasons that are generated within that perspective decidedly subjective. Parent-reasons exist because, often, parents love their children and see their needs and interests as generating reasons to help them. Friend-reasons exist because we care deeply for our friends. Moral-agent-reasons exist because we are concerned for the welfare of those with whom we share lives and communities. Parent-reasons and friend-reasons do not exist because we have moral-agent-reasons; moral agent reasons are not more fundamental in some sense. Being a moral agent carries with it a moral point of view – perhaps one where we view others equally and treat them with deserved respect – but the reasons that we have as moral agents need not come from the value of objects or others or humanity itself. Just as parent-reasons are generated by loving children and friend-reasons are generated by a love for our friends, moral agent reasons are generated by a concern for others. On this view, a distinct moral point of view does not generate a distinct class of objective reasons, but rather a certain kind of subjective reason – one that is generated by a very personal concern for all others. These subjective reasons won't come from viewing the world from an impartial perspective or as a citizen of the Kingdom of Ends, but from considering what's best from our practical identities.

When we reflect on what we have reason to do, we think about what we value, what's important to us. But I deny the move from thinking about that to the requirement that we must, therefore, value our humanity and the humanity of others. We might, as a contingent matter. But

it's not necessary. And so, it shouldn't be a necessary requirement on some consideration's being a reason that someone who values humanity in the right kind of way also would be motivated to this other thing. With this unwelcome revision, Korsgaard's and Nagel's view become decidedly subjectivist. Yet, it is the move they must make if they are to avoid the charge that their arguments beg the question in favor of objectivism.

2.5 Final Thoughts

Objectivism about reasons seems necessary because it is thought that objectivism is true for morality to exist at all. If there are moral facts, it must be the case that we have reasons to act in accordance with them, or that we would be motivated to act in accordance with them if we considered things aright. In this chapter I've put pressure on that assumption. Facts are not reasons unless we assume that that's what reasons are – unless we assume objectivism about reasons. But that's precisely what's at issue here. A fear of the collapse of morality makes this assumption seem inevitably true. But that seeming so is not an argument for objectivism.

At the end of "Morality and the View from Nowhere," Wolf calls for a different conception of morality. She argues that we must mean by morality not the corpus of facts about the properties of objects, rational principles or what's valuable from the moral point of view. If, with Wolf, we are open to a different account of moral domain, then a denial of the existence of objective reasons shouldn't be so scary. In the remainder of my dissertation, I make the case for a subjective understanding of reasons without giving up on morality. Subjective reasons, I argue, are not only philosophically defensible but also, in practice, preferable to objective reasons.

Chapter 3: Why Subjectivism?

I argued in the previous chapter that there is no non-question-begging way to defend objectivism about reasons. The arguments provided by Parfit, Scanlon, Nagel, and Korsgaard provide support for the view, but at bottom, objectivists must assume that "being a reason" is a universal property of moral facts or values (in either a mind-independent or mind-dependent way). That's simply what reasons are, what they must be in order to morality to exist at all. In this chapter, I'll provide some support for the other side: subjectivism about reasons. I'll review five arguments that have been adduced in support of subjectivism. None of the arguments I'll present are irrefutable; the best I can do is show that subjectivism is the *more* plausible alternative to objectivism about reasons. This and remaining chapters are dedicated to just that.

Recall that subjectivism about reasons says that what makes some consideration a reason is the relation that consideration or fact holds to a subject's desires, concerns, pro-attitudes, subjective motivational set or the like. According to the subjectivist, facts about what things are objectively valuable (if there are such facts) are reasons when some subject also antecedently desires or cares about those things. An object's value does not itself generate a reason to pursue, promote, or obtain that object. It's not the value of objects but rather the desires of the subject that make some consideration a reason. Why think subjectivism is right? Five arguments I'll present here are (1) disanalogy to reasons for belief, (2) argument from the aims of action, (3)

evidence of reasons, (4) avoiding presumptuousness, and (5) avoiding alienation. The arguments are presented in order from least to most convincing, at least to my ear.

Though each argument has its limitations, considered together, the five arguments canvassed in this chapter present subjectivism as better accommodating of difference than objectivism and less metaphysically dubious; it has parsimony on its side. Together, these features of subjectivism provide the two most compelling reasons to support subjectivism – the two pillars holding up the view ahead of objectivism.

It's important to note that these are arguments for subjectivism in general; I don't take myself to be arguing in favor of a more specific position with the subjectivist camp. I will distinguish among specific subjectivist positions in later chapters. For now my goal is motivating why subjectivism is an attractive alternative to objectivism.

3.1 Disanalogy to Reasons for Belief

One argument for subjectivism plays up the differences between reasons for belief and reasons for action. Invoking an analogy between reasons for belief and reasons for action, Markovits (2014) claims that the most plausible theories of reasons for belief, externalist accounts of epistemic reasons, have no parallel in the practical case. Markovits observes that epistemic reasons and practical reasons are alike in one important sense: they both count in favor of doing something. Reasons for belief count in favor of believing a proposition, while practical reasons count in favor of doing something. The correct accounts of reasons for belief, then, might provide a clue to the correct account of reasons for action. Markovits argues that, actually, the correct account of reasons for belief – an externalist account – has no parallel in the practical

realm. As a result, the correct account of reasons for action is not an externalist (or objectivist) account.

According to internalist theories of reasons for belief, what we have reason to believe depends "only on what we already believe and the standards of procedural rationality" (Markovits, 2014, p. 60). In other words, we believe in accordance with our evidence and aspire to have maximally internally consistent sets of beliefs. Internal reasons justify a belief in virtue of the relationship they show the belief to stand in to our other beliefs. So, for example, if I believe that the world is flat, I have an internal reason to believe that Christopher Columbus's ship will "fall off the face of the earth" when he reaches earth's end, and no reason to believe that his ship will come back around from the other side. But of course this is the major problem with internalism about reasons for belief. It seems possible to have an entirely consistent set of beliefs that is nonetheless, "wildly deluded," to borrow Markovits term (p. 60). Internalism about epistemic justification, which says that all beliefs justified in virtue of the relationships they stand in to our other beliefs, must be wrong. And so, it seems, reasons for some beliefs must exist externally to what we already believe. They have another source: the world around us.

External reasons count in favor of believing a proposition P not in virtue of the relationship P stands in to our antecedent beliefs, but rather in virtue of the relationship P stands in to what the world is really like. We know what the world is really like by experience, and regardless of what we currently believe, experience alone seems to give us some reasons to believe certain basic propositions. Christopher Columbus's direct experience provides reason for him to believe that the earth is not flat*,* even though this new belief stands at odds with the rest of his beliefs. The fact that I have an experience of redness gives me a reason to believe that I'm having an experience of redness regardless of what else I believe.

So-called foundationalists about epistemic reasons champion externalism about epistemic reasons, at least with respect to a certain class of beliefs, namely those that are basic.[10] According to foundationalism, basic beliefs are justified directly and non-inferentially, such as by sensory experience. We need not infer that we ought to believe proposition P by considering its relation to other beliefs, we believe it directly because we recognize some unlikely-to-be-false reason to do so. Foundationalists agree that there must be some reasons that do not depend on the having of other justified beliefs to count in favor of forming beliefs. For example, at least (some) sensory experiences can provide externalist reasons to form new beliefs that are unrelated, or even contradictory, to current beliefs. If this is so, Markovits argues, it is because sensory experience is an uncontroversial source of information about the world - one that is largely immune to human error. At least some beliefs – basic, foundationalist beliefs - then, are justified by external reasons rather than internal reasons, namely those basic beliefs anchored in sensory experience.

Next, Markovits asks, is there an analog to basic beliefs in the practical case? We must believe basic beliefs, whatever else we may believe, desire, or want because they are anchored in uncontroversial and incontrovertible source of justification: sensory experience. Thus, Markovits' question queries whether there is an uncontroversial source of information, largely immune to human error, which tells us what we ought to do or ought to want? Is there some action we must perform or end we must adopt, whatever else we may want? Markovits says no: "there's no consensus among philosophers on a reliable means of directly forming simple, uncontroversial, unlikely-to-be-mistaken aims and intentions" (2014, p. 64). She says in a footnote that not even

[10] Some foundationalists views allow that non-basic beliefs are justified internally, rather than externally, and so champion internalism about at least some reasons, namely non-basic beliefs. These more nuanced foundationalist theories are discussed in more detail below.

moral intuitions could play this role, as there is no uncontroversial way to form convictions about what is of value or what gives us reason to act.

Markovits argument has been criticized at various points (Ebels-Duggan, 2015). First, she fails to provide a general account of sources of externalist reasons for belief, and many think that these encompass more than just sensory experience. Markovits gestures at "more permissive" foundationalist accounts of basic beliefs, but quickly dismisses them. But one critic says in response: "Once we move to more permissive epistemic views it is less clear either that we can identify the sort of consensus about what sorts of considerations can figure as externalist reasons for belief, which Markovits finds lacking in the practical case, or that if we could, this would not have some practical analog" (Ebels-Duggan, 2015, ₽ 10). Markovits pins her argument on a highly contentious and extreme version of foundationalism, and she may not get the result she wants (i.e. that there is no practical analog to external foundationalist reasons for belief) with a more permissive account.

I agree with critics on this point. Markovits' claim that there is no practical analog to "basic beliefs" in the practical case, such as "basic actions" or "basic aims" hinges on a narrow conception of basic beliefs as only those beliefs rooted in sensory experience. But on some accounts of epistemic reasons, basic beliefs, or those that are non-inferentially justified, encompass a much wider variety of beliefs. For example, some foundationalists allow that memory can justify basic, non-inferentially justified beliefs (A.I. Goldman, 1979) and others argue that the fact that I find myself believing some proposition P is a prima facie, non-inferential justification for believing proposition P (Huemer, 2001). Once we allow a wider class of entities to justify "basic" beliefs, it becomes less clear that there is consensus that non-inferential knowledge is justified by only external reasons. "I find myself believing P" is an

internal reason to believe P, but counts as foundational if, as one agrees with Huemer, it is nonetheless a non-inferential justification for believing proposition P.[11] Furthermore, there is an obvious practical analog to "I find myself believing that P," namely, "I find myself desiring that X." This is all to say, Markovits' argument from the disanalogy between epistemic and practical reasons may do no work as a justification for subjectivism, since there may be no disanalogy after all, even if one accepts a foundationalist account of reasons for belief.

A second and third objection to Markovits' view counters her claim that, "there's no consensus among philosophers on a reliable means of directly forming simple, uncontroversial, unlikely-to-be-mistaken aims and intentions" (64). Two different objections can be formulated against this proposition. Objectivists about reasons might agree with Markovits that while there's little consensus on a reliable means of forming the right intentions, there is nonetheless a truth of the matter as to what aims and intentions we ought to pursue. Objectivists will assert that we have reasons to perform certain actions or adopt certain ends even if there is no uncontroversial way of discovering what those aims and intentions are. Then again, what is needed to defend a theory of reasons is not only establishing that we have good reasons to believe that moral facts, or facts about what's valuable exist, but also that those facts are reasons to act. Certain beliefs may be inescapable, but the move from belief to intention-formation and action is not. Objectivists, even if they persist in hanging on to the idea that there is widespread consensus or facts about what aims and intention are basic to morality or to human flourishing,

[11] One may be a more traditional foundationalist about belief and deny that "I find myself believing P" or "I remember P" is a justification for believing P. If that's the case, then there is a smaller evidence base for basic beliefs. This strengthens Markovits' argument. Markovits argument is most successful when we assume that the body of evidence for "basic beliefs" consists of only external reasons. She argues that there is no analog to "basic beliefs" in the practical domain, insofar as there is no body of evidence that could non-inferentially justify "basic aims" or "basic actions." Her argument is most convincing when we assume that only external reasons could justify basic beliefs. Once we let in internal reasons for even basic beliefs, the disanalogy between epistemic reasons and practical reasons disappears.

There are two possible responses open. First, Peggy's pleading with Jane can be interpreted as a process of reasoning with her, of trying to convince Jane to order her preferences differently "by her own lights" rather than interpreted as convincing her that there is some objective reason to prioritize a friend's (or stranger's) needs above one's subjective preferences. This interpretation of their exchange is entirely consistent with subjectivism about reasons. Second, insisting that Jane has a reason to put Peggy's needs ahead of her own desires and life goals does not amount to a defense of the claim that she in fact has a reason to do so. Peggy may wish that Jane would stay and care for her, or want her to. Many others may wish the same. But wishing it is the case does not establish that it is so. In certain cases, like Peggy and Jane's, it is not presumptuous to assert one's wishes, but one must also acknowledge that the world may not conform to how one wants it to be. Jane may not have a reason to put Peggy's needs above her own desire to snowboard, no matter how much we plead with her.

3.5 Avoiding Alienation

Given that subjectivism starts with the presumption that everyone's desires, concerns, cares, ends, and other elements of her subjective motivational set matter and are equally worthy of pursuit, it avoids alienating individuals from their reasons. Alienation from one's reasons occurs when what one has reason to do lacks a connection to what she would find in some degree compelling or attractive, at least if she were reasoning correctly and sufficiently informed. Alienation from our (objective) reasons is an almost inevitable result of objectivist theories. When one's reasons bear no connection to her subjective desires, goals, or ends, we can expect that acting in accordance with those reasons might contradict or thwart one's desires, goals, or ends. If all normative facts have the property of being a reason by definition, as they do

according to Parfit and Scanlon, then whether or not those facts matter to someone, motivate her, or satisfy her desires is irrelevant to its being a reason. Similarly, if reasons are interpreted as O-I reasons instead, as what someone would be motivated to do from an impartial or perfectly rational perspective as Nagel and Korsgaard insist, then what our subjective and particular concerns are is of little or no importance to establishing what we have (objective) reason to do.[15] Acting in accordance with objective reasons entails, then, alienation from our subjective and particular concerns in favor of those facts or concerns that matter to some, as Railton puts it, "arbitrarily different rational being" (Railton, 1986a, p. 9).

Subjectivists and objectivists can agree that something has gone wrong when a person is alienated from their reasons. They are both "against alienation" in some sense. They disagree, rather, on whether the conundrum that alienation presents is a result of an incorrect theory of reasons or a morally perverse person. Tiffany (2003) argues that when it is pointed out that persons are alienated from their reasons, subjectivists place blame with the theory, insisting that we should reject a theory of reasons that alienates persons from their reasons. In contrast, when alienation is point out to objectivists, they fault persons, insisting that if the person was really viewing the situation impartially or thought harder, she would come to realize that she does have a reason, thereby vindicating an objectivist account. There is nothing in the concept of alienation itself which could indicate which of these interpretations of the problem it produces is correct.

Intuitions about when some individual is problematically alienated from her reasons depend on the conception of reasons adopted by the intuiter. Alienation is a problem for

[15] Korsgaard might try to reject this. She might say, for example, that our subjective concerns and practical identities are necessary, as we can only reason backwards from our practical identities to the conclusion that our moral identities place certain demands on our acting, e.g. acting in ways consistent with the Formula of Humanity. Still, the content of our subjective concerns and practical identities is of little to no importance to determining what the Formula of Humanity demands of us, which is to say, of little to no importance to determining what we would be motivated to do if we were reasoning correctly from the perspective of a citizen in the kingdom of ends. So, I believe this assessment of her view is correct.

subjectivists when supposed reasons for action bear no relation to what the individual desires or cares about. On the other hand, alienation is a problem for objectivists when the cares, concerns, or motivations of an alienated individual do not cohere with her (objective) reasons. Again, both sides agree that alienation is a problem, so the fact of alienation does not, according to Tiffany, tip the scales either way.

While I agree with Tiffany's analysis of the way subjectivists and objectivists appeal to alienation in the debate over reasons, I do not agree with his conclusion that the argument from alienation does not tip the scales one way or the other. On the contrary, I think it does, and in favor of subjectivism, for similar reasons mentioned above in the argument against presumptuousness. We should start with the assumption that everyone's ends matter and are worthy of pursuit, not that only certain ends are worthy of pursuit (e.g. those that are consistent with universal values or principles or obtain from a moral point of view). When faced with the problem of alienation, we have the option of blaming persons (for their inability to see that universal values or moral facts constitute reasons) or blaming theories (for failing to take into account that ends vary, and the moral perspective is one among many, equally valuable ways of perceiving the world). When faced with whether to place blame on a theory or on a person, I think we should side with people, revising the theory to better accommodate individual differences that give rise to large-scale alienation from supposedly objective reasons.

Even on a subjectivist theory of reasons, there will be cases where alienation from one's reasons is inevitable. I have in mind cases of irremediable misinformation, mental illness, unwilling addiction, or brainwashing. In these extreme cases, a person may be alienated from her reasons, insofar as the expressed desires or motivations, e.g. craving a cigarette, conflict with what she would want in improved circumstances, freedom from symptoms of addiction. (This is

not to imply that willing addicts can have reasons to continue pursuing drugs.) But these cases are not like most.

The interpretation of the problem of alienation from a subjectivist perspective also has parsimony on its side. Locating in the problem of alienation in the person rather than the theory of reasons requires the objectivist assumption that reasons exist in some metaphysically vague way, as I argued in Chapter Two. On the other hand, locating the problem of alienation in an (objectivist) theory of reasons rather the person requires only the non-controversial existence of subjective desires, concerns, motivations, pro-attitudes. The truth conditions on objectivist reasons require the assumption of metaphysically "queer" facts and dubious assumptions about human psychology, while the truth conditions on subjectivist reasons can be met by positing the existence of only non-controversially extant entities. An argument from parsimony favors the subjectivist interpretation of the problem of alienation. Where both sides of the debate recognize that alienation should be avoided, only one side has a naturalist solution.

Conclusion: The Pillars of Subjectivism

The five arguments canvassed in this chapter have been adduced in favor of subjectivism about reasons. They are not irrefutable arguments; on the contrary, I readily admitted that objectivist responses and/or other kinds of objections are available to those who reject subjectivism about reasons. Notwithstanding these limitations and objections, I think that the five arguments present subjectivism as, on the whole, less metaphysically dubious and better accommodating of personal differences than objectivism. These are the two most compelling reasons to support subjectivism.

The arguments against presumptuousness and against alienation, and the argument from evidence of reasons show vividly, I think, that subjectivism accounts for inevitable and vast differences in what is valuable to particular individuals better than objectivism does. In stipulating *a priori* that certain ends or values constitute reasons for action and allowing some (maybe most) individuals will be alienated from (objective) reasons, objectivists ignore cultural and personal differences (that effect differences in what people have reason to do) in itself an objectionable way. I've often heard statements like, "You're in a different situation – what should *I* do?" or "I know you should do that, but what should *I* do?" or "But I don't want the same things as you!" These statements all support the notion that who I am is inextricably linked to normative reasons. Objectivism cannot accommodate those kinds of statements. O-E will insist that whatever else we want, there are certain things that we must have reasons to pursue or desire. O-I will say that whatever partial perspectives or practical identities we embody, we must realize that there is one perspective we share – the impartial perspective or a moral identity – and that we all have the same reasons in virtue of sharing that perspective. I'm not convinced that we must share that perspective, nor am I convinced that it is desirable that we do. The argument from evidence for a reason, and the arguments against presumptuousness and against alienation are compelling because they present subjectivism as better accommodating difference.

Furthermore, subjectivism is presented throughout as less metaphysically dubious. It does not require positing "queer" metaphysical facts or developing a complicated (and arguably incorrect) model of human psychology and rationality. As seen in several of the arguments, objectivists can agree that for some reasons, e.g. reasons to choose a restaurant, subjectivism is correct. Our practical identities and contingent desires often do generate reasons. Objectivists go on to say that's not the whole story. There is another class of reasons, namely objectivist ones,

like the theories reviewed in Chapter Two, which captures another aspect of normativity. But why do we need both? If subjectivism can be constructed in a way that accommodates our moral intuitions and preserves important aspects of normativity and moral experience, why multiply concepts and metaphysical kinds? Parsimony demands that we avoid doing so. Subjectivism about reasons is consistent with naturalism about normativity. The alternative view requires non-natural properties and a special sort of intuition or insight to discern them. The subjectivist move seems the only alternative to treating the property of "being a reason" as a further non-natural property whose perception is itself the object of some special sort of cognitive act.

Chapter 4: Ideal Advisor Accounts of Reasons

In the previous chapter, I provided five arguments in favor of subjectivism. The two most compelling reasons to support subjectivism, I argued, is that it is consistent with a naturalistic metaphysics, and vindicates the idea that what we have reason to do depends on who we are, what we want, and what we are concerned about.

But subjectivism cannot be the view that we have reason to do whatever we happen to want to do. Bernard Williams' infamous gin case captures this point (Williams 1981). In the case, a man desires to have a gin-and-tonic. There is a glass in front of him which he believes contains gin-and-tonic, but which in fact contains petrol. In this case, Williams says, it is incorrect to attribute to the man a normative reason to drink the contents of the glass in front of him, even if it reasonable for him to drink it, given that he believes it contains gin. Since there is no gin in the glass, however, he lacks a reason to drink it. What we have reason to do isn't grounded in any old desires, but rather only those desires that are suitably informed. In other words, we undergo an "idealization process" to correct misinformation, false beliefs, or faulty reasoning that leads to our having a particular desire. The desires that remain, or the desires that you would want yourself to have, after this process are the desires that ground true claims about what we have reason to do. The idealization process is an attempt to get non-normative existing facts into clear view, e.g. facts about what it would be like to Φ rather than to Ω or facts about one's set of subjective motivational set. So, subjectivism must be revised off the bat. It doesn't say that all

desires generate reasons, but rather that those desires that survive the idealization process generate reasons.

What the idealization process looks like varies from subjectivist account to subjectivist account. In this chapter, I will present a particular subjectivist account of reasons, namely the Ideal Advisor Account (IAA) of reasons. IAA says that A has a reason to Φ when a fully informed version of herself, A+, would want her non-ideal self, A, to want to Φ in her actual circumstances given A's desires, concerns, cares, and the like. A's desires ground what makes something a reason for her, but the process of becoming fully informed ensures that A's reasons are grounded in full information about the objects of her desires. IAA falls into the subjectivist-*externalist* camp. That is, it is a subjectivist account of reasons because it says that reasons are grounded in agents' desires, but it does not make the further claim that reasons would serve as the motives for particular actions. In other words, an agent can have a reason or action even if it would not motivate her action under any circumstances. The guiding idea behind these accounts is that if agents were in improved circumstances, e.g. if they had full information, their desires, and so their reasons, would be different. The differences between desires and ends of fully informed agents and their less-informed counterparts explain why less informed agents have reasons even when those reasons could not explain or serve as the motive of action in present circumstance. Externalists are open to the possibility that normative reasons sometimes are our basis for action, but they deny that it *must* be the case that our reasons serve as the basis for our action, i.e. that our reasons would motivate us to act.

After raising several objections already found in the literature on IAA, I will raise a novel objection, arguing that the IAA leads to results that undermine the theory's subjectivist commitments, insofar as A+ may recommend that A have desires utterly alien to her current

motivational set. IAA leaves open the possibility that A+ will recommend that A has a reason to act contrary to her deeply held convictions and beliefs - those elements of her subjective motivational set, or S, unique to and constitutive of her identity. As such, IAA permits cases where A is alienated from her reasons. I will suggest that a more moderate version of A+ should be spelled out to produce a more satisfying subjectivist account of reasons. In subsequent chapters, I go on to defend subjectivist-*internalism*, inspired by Bernard Williams' influential account of internal reasons.

4.1 Ideal Advisor Accounts of Well-Being and Reasons

So-called Ideal Advisor[16] accounts of well-being hold – very generally – that an agent's life goes best if she pursues those things that she would want if she had full knowledge of the causal outcomes of the actions available to her. In other words, certain actions or objects constitute our good when we would choose them if we knew all of the options available and the consequences of choosing those options. When we do not know all of the options or means available to satisfy our concerns and desires, we regularly choose to do things that are not in our best interest. We often do not have enough information to make the best choice. But that does not vitiate the "best choice," it entails we will (sometimes) be incapable of choosing it since we lack information and/or the means of obtaining more information. To use an example, let's say scientists find out in fifty years that decaffeinated coffee is bad for you in light of its carcinogenic effects. If this were to happen, it would be true that decaf coffee is contraindicated for those people who want to avoid cancer. It would be true even now that it is in the best interest of decaf coffee drinkers to stop drinking it, even though they lack access to that information.

[16] "Ideal Advisor" accounts of the good and of reasons are so-called due to their structural similarity to earlier "Ideal Observer" analyses of ethical concepts. See Frith (1952).

The idea that what is good for us is what we'd determine is best if we knew more information has a long history in philosophy. Mill's claim that "impartial judges" should be consulted to determine what we ought to do is an early model. We get from Mill the idea that if a person prefers ϕ-ing to Ω-ing while fully acquainted with both "in a cool hour," then ϕ-ing is more conducive to the agent's well-being (Mill, 1861/1979, Chap. 2). Sidgwick offered the first explicit formulation of the account, saying:

> A man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realized in imagination at the present point in time (Sidgwick, 1981, pp. 111-2).

Harsanyi, Brandt, Hare, Rawls, Guathier, Griffin, Darwall, Lewis, Railton later developed or endorsed some version of the view.[17]

Sidgwick's initial formulation of the view (and those like it) quickly runs into difficulties. A fully informed person would never want more information for herself, but we are firmly convinced that it is often in the best interest of uninformed individuals to do some research. For example, Sobel (2001b) says, "The idealization process turns us into such different creatures that it would be surprising if the well-being of the two of us, my informed self and my ordinary self, consisted in the same things" (p. 477). The presence of a desire in the ideally informed agent does not give us grounds to suppose that satisfying this desire would be good for the non-

[17] Brandt suggests that we identify what is good with what a person would "rationally desire" after maximal confrontation with facts and logic (Brandt, 1979, p. 1). R.M. Hare adopts an informed preference account akin to Brandt (Hare, 1981). Rawls identifies a person's good with his "rational plans of life…that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all of the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires" (Rawls, 1971, p. 417). Darwall endorses Rawls's view (1983). For related views see Harsanyi (1977), Gauthier (1986), Griffin 1986, and Lewis (1989). Harsanyi, Gauthier, Griffin, and Lewis each develop an account of "full information" that is differs slightly from each other in various ways. I consider Railton's account, and Sobel's nearly identical formulation of it, to be the canonical version of Ideal Advisor accounts of reasons since it avoids the difficulties mentioned in the text. Smith (1994) defends an account of reasons similar to Ideal Advisor accounts. I will ignore Smith for the purposes of this chapter, as I consider him an objectivist about reasons, insofar as he builds substantive evaluative judgments and desires – a particular moral outlook – into his account to the Ideal Advisor. He argues that all individuals have the same (moral) reasons, insofar as the desires of our Ideal Advisors converge whatever else our actual selves want.

idealized agent, and vice versa. The difference in information and outlook of the ideally informed and non-ideal agent amount to a difference in what they have reason to do. So, we cannot look to the ideally informed agent's desiderative profile to fix what the non-ideal agent has reason to do.

To avoid this problem, Railton adopted a "wanting to want" or *advice* model of IAA about well-being. Rather than saying your good is constituted by want you would want if you were fully informed (which, we expect, differs from what you want right now), Railton says your good is determined by what your ideally informed self would want your non-ideal self to want in present circumstances. To that end, Railton says,

> An individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses in instrumental rationality (Railton, 1986a, p. 14)

What does it mean to contemplate a present situation from a standpoint fully informed about himself and his circumstances? Railton suggests a quite technical procedure, termed the "reduction basis." What renders some end or activity part of an individual's good is not the fact that he would want himself to want it were he fully informed, but rather the existence of the reduction basis for that counterfactual, namely, "the particular constellation of law-governed features of the actual individual and his circumstances in virtue of which these claims about idealized hypothetical desires hold" (ibid., p. 25). The "law-governed features of the actual individual" include her pro-attitudes, desires, concerns, interests, ends, etc., as well as her circumstances. These are an individual's "non-belief properties" (ibid., 20). Knowledge of these features is supplemented with general knowledge about the world, including, for example, perfect knowledge of what would happen if one ϕ's instead of Ω's. In the process of becoming ideally informed, we cannot change one's pro-attitudes, desires, concerns, interests, ends, etc. in

a way that violates psychological laws. (Railton does not specify which psychological laws he has in mind.) Adding in that the idealization process is law-governed ensures that the is-ought distinction is maintained. It will not be the case that someone ought to do what is psychologically impossible given who she is and her current desiderative profile.

Conceptualizing the idealization process as a reduction basis likens the process of becoming ideally informed, and then recommending a particular course of action to one's non-ideal self, to a complex algorithm. We input the non-ideal self's desiderative profile, subjective motivational set, her abilities and circumstances. We add in general facts about the world and knowledge of the effects of available courses of action. The output is a singular recommendation of is best – an "objectified subjective interest" in Railton's terminology (1986b, p. 173). The recommendation is subjective insofar as it is rooted in the subjective desires, ends, or concerns of a particular individual, but it establishes an (objective) fact of the matter as to what is *actually* in her best interest, as opposed to what she may think is in her best interest from the non-ideal standpoint. The reduction basis is designed to determine definitively what's in the best interest of the non-ideal self. Railton's view is, therefore, shielded from the worry that "what one would want her non-ideal self to want" is just a matter of idyll musing. Railton admits that the Ideal Advisor, so called A+, is not a person in any real sense, but rather represents what a fully and vividly informed version of you, free from cognitive errors and lapses in instrumental rationality, would want.

Those who are convinced by IAA when it comes to well-being rely on similar considerations and arguments to defend IAA about reasons. Just as what's in our best interests is determined by fully informed versions of ourselves on IAA about well-being, defenders of IAA about reasons (esp. Sobel, 2001a, 2001b) argue that our reasons are fixed by the

recommendations of a fully informed version of ourselves, A+. A+ has full information about A's "non-belief properties" and circumstances, is fully rational in the sense she suffers no cognitive defects, she has access to general knowledge about the world and the courses of action available to A, and makes no mistakes in instrumental reasoning. The rationale for granting the idealized agent this information and experience is, "to provide her with a more accurate understanding of what the options her non-ideal self is considering would really be like" (Sobel, 2009, p. 343). It is important to notice that while this ideal standpoint builds into the idealization process a lot of new information, it changes "non-belief properties" of A – her desires, preferences, concerns, goals – only to the extent that newly acquired information about the world and the consequences of possible courses of action affects her desires, preferences, concerns, goals, and other elements of her subjective motivational set. In this way, Sobel's suggestion respects Railton's reduction basis used to determine one's objectified subjective interest. Any changes in A's desires, concerns, ends, personality traits, etc. must be consistent with laws that govern psychology, and in particular, changes to those kinds of psychological phenomena. But it retains the subjectivist insight that an agent's current subjective motivational set ground what makes some consideration a reason for her.

4.2 An Externalist Account

Sobel, one of the biggest champions of subjectivism in contemporary metaethics says bluntly, "the best subjectivist accounts of reasons for action must tolerate external reasons," and "it is not a necessary condition on consideration C providing A a reason to X that there be any particular version of A that can conclude via a sounds deliberative route that he ought to X" (2001a, p. 233). In other words, it is possible that neither A nor A+ would be motivated to act on

the basis of consideration C, even though C is a reason for A. Sobel's externalism echoes Railton's:

> A fully informed and rational individual would, or example, have no use or desire for psychological strategies suited to circumstances of limited knowledge and rationality [e.g. a reason to get more information]; but he no doubt would want his incompletely informed and imperfectly rational actual self to develop and deploy such strategies" (Railton, 1986a, p. 16).

A+ wants her non-ideal self to develop and deploy such strategies even if A would not be motivated to do so given A's lack of information or her spatiotemporal location. Thus, while A would not be motivated to act on the basis of a reason he has given the impossibility of becoming aware of that reason, A+ lacks the reason altogether. So, the reason will not serve as the basis for action in any possible world.

Sobel calls these kinds of reasons "fragile" reasons. He says, "A's reason to Φ is fragile if and only if A has it but A+ lacks it" (Sobel, 2001a, 231). Fragile reasons are Sobel's key support for thinking that internalism about reasons is misguided. So long as fragile reasons exist, internalism is wrong. Of course all Sobel needs is one counterexample. He provides the example of a singular taste. A food with a singular taste is one such that once one has tasted it, one is glad to have done so, but never wants to taste again. Assuming that A has never tasted this food, A+ will want A to try it. But having already tasted it (or at least come to know the taste in the idealization process), A+ does not want to try it, and so has no reason to do so.

Sobel thinks A's reason to try the food with singular taste as not only a fragile reason (one which A has but A+ lacks), but also a superfragile reason – a reason "so fragile that the only vantage points from which one could appreciate the way in which Φ-ing furthers something in the actual agent's subjective motivational set are vantage points in which one lacks a reason to" (ibid.). In sum, "Superfragile reasons are reasons that one cannot have and be motivated by

simultaneously" (ibid.). The possibility of fragile and supergragile reasons are the central argument for externalism about reasons while defending subjectivism.[18]

4.3 Objections in the Literature

IAA has a lot going for it. First, it is consistent with a thoroughgoing naturalism and the parsimonious metaphysics as discussed in the previous chapter – a compelling reason to accept any subjectivist view of reasons. Railton (1986a, 1986b) indeed defends his account in part with an appeal to its naturalism. Ideal Advisor views are naturalistic insofar as it provides a purely descriptive analysis of a person's good, one that retains normative authority of the good, but does not incorporate substantive evaluative judgments or universal concepts of the good or of reasons. Second, IAA gets support from several common intuitions. It makes sense of the common observation that experimenting with alternatives frequently increases the agent's knowledge of her good. That is, we typically think that we become better at determining what is good for us as we gain more experience – we become older *and* wiser. We seek advice from people who have experienced various courses of action when we are trying to determine what to do since we recognize that our own ability to choose the right action is limited by our lack of information and experience. Finally, it respects the seemingly powerful thought that if a person cannot be brought to value something, then it is not valuable to her.

[18] The possibility of fragile and superfragile reasons is challenged in Robertson (2003). Robertson's case against the existence of fragile reasons asks: why must A be motivated only by the taste of the food? If A is the kind of person who likes to try new and odd foods, or is very curious about new experiences, she may be motivated to try the singular food even though the taste alone is not what is doing the motivating. Her more informed self will recommend that A try it, knowing that it would satisfy A's curiosity and desire. In this case, A has the reason and is motivated by it, contrary to what Sobel insists must be the case. I find Robertson's objection successful, but will not go into more detail here. I think externalism about reasons is objectionable for other reasons developed in sections 4.3 and 4.4.

Yet for all those positives, IAA is vulnerable to objections. Broadly speaking, there are two lines of attack of Ideal Advisor accounts of the good and of reasons. The first says that the Ideal Advisor is a fiction, as it is impossible for a being like us to achieve full information or to make recommendations in the way stipulated by the theory. The second says that the Ideal Advisor lacks the normative authority to determine the reasons of her non-ideal self. In other words: why should our non-ideal selves be subject to the recommendations of a possibly very different person? This second way of attacking Ideal Advisor accounts grants that the Ideal Advisor may be able to provide a determinate answer to the question, "What do I have reason to do?" but it provides reason to doubt that the Ideal Advisor's answer is normatively binding. I'll present both types of objections, and then flesh out a novel objection to IAA in the next section.

4.3.1 Ideal Advisor as Nomologically Impossible

In his earliest paper on ideal advisor accounts of well-being, Railton owns up to the following objection: "there are grounds for skepticism about whether there is such a thing as 'what one would, if fully and vividly informed and fully instrumental rational, want oneself to seek were one to assume one's actual place'" (1986a, p. 19). More recent papers have elaborated on the grounds for such skepticism. I'll explain three grounds for skepticism found in the literature, and then add my own worry.

First, because of what it is like to be a person and occupy a perspective, it appears that no human being could be fully informed. Superhuman capacities are required of A to become A+ (Rosati, 1995). Rosati spells out what she takes Railton to mean when he says that our idealized selves are our actual selves but with, "unqualified cognitive and imaginative powers" (Railton, 1986a, p. 173). Rosati says of this person:

> At a minimum, she would have to have capacities of reason, memory, and imagination far surpassing those she actually has. She would have to be able to have all of the necessary experiences and keep them clearly before her mind, remembering them as experienced in themselves and as experienced in relation to what comes before and after. In addition, she would have to retain features of her personality that enable her to experience her lives as she would as the persons living them, desiring and being motivated as she would from within those lives, while losing all features of her personality that keep her from absorbing information. At the end of the process of imaginatively surveying her possible lived in all their permutations, she must have…traits that enable her to appreciate what each experience was like, rather than traits that enable her to appreciate only some experience while not appreciating others" (Rosati, 1995, p. 310).

The main point of this objection is that a person who can do all of this is no longer even subject to psychological laws or limitations of human cognitive abilities. Human psychology and cognition is not "unqualified." What kind of creature do we turn into when we gain those powers? Arguably, some kind of superhuman computer, capable of feeling and thinking as humans, which is to say, capable of experiencing world in the same way that we do, yet with unlimited powers of memory and nuanced comparison. We are humans no more.

Second, effects of the order of information presented to A in the process of becoming A+ will skew A+'s recommendations (Sobel, 1994). Railton (1986a) acknowledged this problem, saying, "the possible effects upon the individual of the *order* and *mode* of presentation of the information he receives create difficulties" for a view according to which A+ makes determinate recommendations after having considered the variety of possible outcomes and lives for the non-ideal A (p. 21). Sobel (1994) considers various ways in which we could construe A+'s process of coming to be fully informed about A's possible actions and the consequences of those actions – in effect, A's possible lives. One way in which A+ might become fully informed is serially. On the "serial" version, our idealized self achieves full information by acquiring first-hand knowledge of what it would be like for A to choose to ϕ, then what it would be like for A to

choose to Ω, then Ψ, and so on and so forth.[19] The problem is that as A+ experiences these different activities and courses of action, what she believes, desires, and values within one possible course will impact what she comes to believe, desire, and value in the next. It could interfere with her ability to choose impartially which one is, in fact, best for A, that is, to determine which action(s) A actually has reason to pursue. Sobel asks us to consider someone – A – who is considering who to kiss for her first kiss. If the serial model is correct, A+ would need to experience serially A's possible first kisses. But to do that would be to undermine the experience of each first kiss, on each kiss after the first first kiss. This is a somewhat silly example, but it gets at the point motivating this concern with ideal advisor accounts of reasons: past experiences shape future experiences. In demanding that A+ experience fully the nuances and phenomenology of A's various possible lives, we demand that she do incompatible things. We demand that she experience each life as novel and fresh, but certain experiences radically affect one's perspective. So, this way of construing the idealization process cannot be right.

Another way of idealizing A is through the "amnesia" version. According to the "amnesia" version, A+ experiences what it would be like for A to choose to ϕ, then what it would be like for A to Ω, then Ψ, and so on and so forth. But between each possible life, A+ forgets the experiences, beliefs, and values of the previously lived life. A+ experiences each discrete life, and then is allows epistemic access to each, as if watching each possible course of action back in a movie. A+ then recommends to A one of the possible courses of action – thereby establishing A's reasons. Sobel argues that the amnesic version of A+ is more plausible than the serial version, and we should understand the Ideal Advisor in this way. Though this may seem

[19] We assume that A+ needs first-hand experience of each possible action, outcome, and life if we are to be rationally persuaded that A+ is an ideal position to choose between possible actions. In order to claim that one knows what it is like to perform an action and appreciate its consequences for one's life, one must experience what it would be like to be in those shoes. See Sobel (1994).

plausible at first glance, it is speculative to suppose that there is a single perspective from which A+ can make a recommendation for A. This brings me to the next criticism.

Third, there may be no single perspective from which A+ can make a recommendation, given that, like A, A+ also has varying concerns, desires, and cares *for* A. A+ may be concerned about A as professor, as spouse, as parent, etc., and so we should be skeptical that A+ can make singular recommendations from a unified perspective (Sobel, 1994). This can be called the problem of too many voices. Within and across possible courses of action, individuals take up a variety of roles. Our evaluative perspective changes over time and across those roles even contemporaneously. Why think that the idealized version of oneself is immune to such internal conflict over which evaluative perspective to prioritize at any given point? Even if the "amnesic" version of become fully informed were correct, it defies human experience to think that there is singular perspective from which A+ can provide a recommendation for A.

A fourth worry not mentioned in the literature so far builds upon this third concern. The third worry is a general worry about "too many voices." This is a problem common to all subjective accounts of reasons: how do we unify the variable and variegated contents of one's subjective motivational set, sometimes conflicting roles, and possible course of action into one all things considered judgment? I'll make this problem more precise, however, and show that it leads to a particular problem for IAA accounts of reasons.

The Ideal Advisor knows not only the immediate consequences of choosing to ϕ rather than Ω, but also the long-term consequences of doing so. Think about being at a crossroads. Alex is unhappy in his relationship, and considering whether to leave Robin. Alex has two choices: leave Robin or stay. Alex+ knows not only what it will be like tomorrow to leave and to stay, but what it will be like 10, 15, 30, years. This is in line with a stipulation of the theory: Alex's fully

informed counterparts knows exactly what it will be like for Alex to choose one or the other. Alex+ knows, then, that if Alex leaves Robin, he will undergo a "transformative experience" to borrow L.A. Paul's term. L.A. Paul describes transformative experiences as a kind of experience that is both radically new to you and changes you in a deep and fundamental way. After transformative experiences, we tend to care about very different things that we care about pre-experience (Paul, 2014). With Robin, at present, Alex is curmudgeonly, always a little bit grumpy. Alex+ knows that upon leaving Robin, Alex will finally have the freedom to embark upon a world journey to become, at the end, a new person. (Think Elizabeth Gilbert's journey as chronicled in her 2007 memoir, *Eat Pray Love*). Alex+ also knows that if Alex chooses to stay with Robin, he will eventually, after several years, come to be happy in their relationship. They will have a child, something they previously thought was impossible, and it will bring Robin and Alex closer together in a way that Alex could never comprehend from his current standpoint, because it, too, is a transformative experience.

The Ideal Advisor does not face the epistemic limitations that Alex does. Alex has trouble making a choice about whether to leave or stay with his partner because he does not know, indeed he cannot know, what his life will really be like after enacting a choice. But the Ideal Advisor does know. Alex+ has experienced each possible life course as Alex, and is now in a position to recommend one or the other, which would settle what Alex truly has reason to do. Alex+ is supposed to recommend one or the other depending on which course of action is better for Alex, taking into account what Alex is like, his current psychology, what gets him going. But since the long-term consequences of both actions are transformative, there seems to be no principled way to decide which one is better for Alex. That is, since both staying and leaving lead, ultimately, to a very different Alex, how is the Ideal Advisor supposed to choose between

them? There may be nothing about Alex currently that tips the scales either way, leading the Ideal Advisor to make an arbitrary choice, or make a choice based on his own preferences, rather than with Alex in mind. This challenges the idea that it is possible for the Ideal Advisor to make a singular recommendation. It is especially a problem for views that allow that the Ideal Advisor has so much information about the long-term consequences of acting – possibly too much information. Too much information may make it impossible for the Ideal Advisor to make a principled choice about what's best for Alex.

Ideal advisor theorists may not see these four grounds for skepticism about whether the Ideal Advisor can really achieve a single recommendation as a problem for the view. Railton very explicitly says:

> It is not an objection to a counterfactual that it involves hypothesizing circumstances that are, in the actual course of things, nomologically impossible. It would be excessively skeptical to insist that there is no fact of the matter about how our lives would be changed were…our unassisted memory [to be increased] tenfold, yet such hypothetical circumstances would involve violation of laws of physiology" (1986a, p. 24).

Surely, Railton's "reduction basis" for computing the Ideal Advisor's recommendations could build in to its algorithm the capacity to make fully informed comparisons of possible courses of action, even when those courses of action lead to transformed version of one's non-ideal self. The fact that we mere humans with very limited cognitive and imaginative powers cannot conceive of a human like this does not count against the conclusion that there is, nonetheless a fact of the matter about what is best for A given access to limitless information and experiences. The laws and machinations required to get there may not be available to us to work it out, but that does not count against their existence in theory. In sum, proponents of IAA will not be swayed by any objections that only seek to establish that the Ideal Advisor is a fiction. They admit it may well be impossible for a being like us to achieve full information or to make

recommendations in the way stipulated by the theory, but that does not undermine the truth of the account.

4.3.2 Ideal Advisor is Unrecognizable

This second way of attacking Ideal Advisor accounts grants that the Ideal Advisor may be able to provide a determinate answer to the question, "What do I have reason to do?" but it provides reason to doubt that the Ideal Advisor's answer is normatively binding for her non-ideal self. Rosati (1995) formulates this worry as the following: there is no guarantee that A+ is someone to whom we ought to accord authority over our normative reasons. Even when we grant that inaccessible laws of "unqualified cognition and imaginative powers" yield a determinate answer of what our idealized self would recommend to us, today, we may not think that such being is us. Rosati says:

> We will have discovered what "she" [A+] would want with full information, in the sense that it is this very individual who underwent the process of idealization. And we will have learned what some one "like her" would want, in the sense that it was this sort of person who underwent the process. But surely we have not learned what she would want, or even what someone like her would want, in the senses in which we ordinarily understand these terms. Given the changes that a person must undergo to become fully informed, Ideal Advisor views do not guarantee that we are the persons who occupy the ideal standpoint. A person might thus plausibly contend that the fully informed person would not really be her, just as we now often contend that the person who would result from a procedure that changed us in ways that we regard as alien **would not be us** (Rosati, 1995, p. 310-1, emphasis added).

I've emphasized the last phrase in this quotation to highlight a contentious claim. In order to be a veritable objection to IAA, it must assume either one of two statements. Either A and A+ are literally two different people, such that A+ truly would not be us. This type of claim relies on a questionable view of numerical identity. Or A and A+ are so different that A+ and A lack any resemblance to each other, as we might say when we look back on our teenage years and say, "I don't know who that person is!" In saying that, I don't deny that it was me who performed

certain actions in the numerical identity sense, but rather invoke a feeling of estrangement from who that person is. The first claim – that A and A+ are literally different people – is a speculative metaphysical claim. Attempting to defend it is beyond the scope of this dissertation. The second is, I think, what Rosati has in mind in the above quotation, even though she does not explicitly state that assumption.

Rosati argues that it is because of what it is like to be a person and what it is like to occupy a perspective that the Ideal Advisor – who is may not be like a person and who must occupy a different perspective – "would not be us." Her main point, emphasized above (see 4.3.1), is that the Ideal Advisor, purportedly you, may not be someone whose judgments you'd regard as human, much less regard as a version of yourself, given that the Ideal Advisor has superhuman capacities of memory, information processing, and instrumental reasoning. Is this the kind of person we ought to take advice from? Rosati says, "you will not be motivated [to act] upon learning the desires of your fully informed self by the consideration that it is 'you,' but rather it will be 'you' only if you are motivated" (ibid.). The idea is that the Ideal Advisor's recommendation will be viewed as authoritative only if the recommendation is intelligible to the advisor's non-ideal self, only if the non-ideal agent can recognize herself in her advisor.

Proponents of IAA will not be moved by this concern. After all, they will say it doesn't matter whether the recommendation is intelligible to the advisor's non-ideal self. She need not be motivated to act on the basis of the recommendation. Again, failure to grasp or appreciate the advice of one's Ideal Advisor does not vitiate the fact that it nonetheless establishes what one has reason to do.

Still, I think there is something to Rosati's concern that the Ideal Advisor "would not be us." Her rub against IAA accounts of reasons, and S-E more generally, is that what we have

reason to do should be intelligible to us. That our reasons make sense to us is a desideratum of a subjectivist theory.

Under the guise of subjectivism, the possibility of estrangement from one's psychologically continuous Ideal Advisor seems odd, perhaps even impossible. How can we feel alienated from some reason or action when it is best for us *by our own lights*? My answer is that even if the process consisted in the perfect algorithm, since the theory allows that we won't be able to perform (or even understand) the algorithm, it leaves open that we have no clue how or why we supposed to ϕ instead of Ω. I may respond to the recommendation to ϕ: "Sure, some better version of myself wants me to ϕ. But how did *she* get there? How did I turn into *that*, a person who wants me to want *this*? That's not a person I want to be."

Opponents might play down this possibility. Sobel, for example, says, "Agents are typically not alienated from the desires they would have if they were more informed" (Sobel, 2009, p. 351). It is some consolation that if it really is our better selves issuing advice, they will likely want things for us that will move us and make our actual lives worth living. The Ideal Advisor will not have undergone a genuine idealization process unless she takes into account all the facts about what our actual selves are like, our actual psychologies, what is good for us, what gets us going, etc. This ameliorates some worries about the distance between the ideal adviser and her actual self.

Even so, atypical cases constitute counterexamples to the view. Yet, there is nothing in the theory to preclude estrangement from our ideal (or non-ideal selves). The cases may be rare, or weird, but they are possible. I will now turn to an example of such a case. The case exemplifies Rosati's worry that the Ideal Advisor "would not be us." But while Rosati is worried that the Ideal Advisor is unrecognizable insofar as it is "superhuman," I will raise a different

But, insofar as subjectivism is primarily concerned with connected people's desires and their reasons, we should eschew a theory of reasons that leaves open the possibility that A cannot be made to see how or why she has a particular reason. In sum, the idealization process is too onerous. If actual agents cannot perform the calculations of the Ideal Advisor, then we've created too much critical distance between our current subjective motivational set and that of the Ideal Advisor.

The problem is not that it is impossible to achieve a determinate answer, but rather that it's that it will always be impossible for us limited deliberators to perform or even understand the algorithm, which contributes to further alienation from our reasons. Alienating us from our reasons stands at odds with one of the main reasons for supporting subjectivism over objectivism about reasons. We can't enter the perspective, supposedly numerically and psychologically consistent with me today, from which issues these recommendations. We couldn't be brought to see that what our reasons are given the amount of information and computational complexity of arriving at the correct determinate answer to the question "what do I have normative reason to do?"

Conclusion: Motivating Subjectivism-Internalism

IAA runs into trouble by hoping that A+'s recommendations will not alienate A from her deeply held convictions and projects. I hope that the example provided shows that assumption is unwarranted, but other examples can easily be generated to make the same point. To borrow from Manne (2014), I am quite convinced that even if a fully informed version of myself told me that my life would go best if I converted to scientology, and so recommended that I study its

tenets and practices, I would deny vehemently that I have a reason to a purchase a scientology indoctrination pack.

The contents of an agent's subjective motivational set – the result of historical, familial, religious, social, cultural, and political forces – constitute in an important way who that agent is, and constrain, as well, who she will become. This is not to imply we are stuck in our own worlds, or that our reasons are generated only by our actual desires. We can and do gain critical distance through new information, new stories, and new experiences. The perspective that we ought to use to determine what agents have reason to do must have some critical distance (in order for the account to be normative) but must not be so far off as to render those reasons utterly alien to the person to whom those reasons attach. A+ cannot be so distant from A so as to recommend things that are off the table for A. If access to information would motivate A+ to recommend to A doing something so far afield from what A is motivated to do, then A+ is no longer a trusted advisor to A but rather an omniscient observer of A's mistakes.

Proponents of IAA are unlikely to be persuaded by my example and my argument. As externalists, they will be unmoved by the worry that what A+ recommendation are unthinkable from the first person perspective or rooted in alien desires. After all, according to IAA, what A has reason to do need not motivate her actions in any circumstances. We may appear to be at a standstill in the debate, but we are not. In the remaining chapters, I present positive arguments for thinking that we should shorten the critical distance between A and A+, as described above, in order to achieve the result what A has reason to do would be possible for her to act on, insofar as it will be intelligible from a not-so-idealized first personal perspective. This amounts to a positive argument for the so-called "explanation constraint" on reasons, the crux of subjectivist-internalist accounts of reasons.

GI secures two theoretical desiderata: (1) any account of what reasons are must make plain how, roughly, *the* reason there is for A to ϕ could be *A's* reason for ϕ-ing. In other words, it must secure a connection between the reason what A is, or would be, motivated to do under suitably idealized conditions. I call this the explanation constraint on reasons. And (2) something could be A's reason for ϕ-ing only if it is also a justification for ϕ-ing. ϕ-ing is justified on GI because the reason issues from a suitably idealized agent, not her actual self. This is the normative constraint on reasons.[24]

5.2 The Conditional Fallacy

Unfortunately the normative constraint that makes GI, and not NI, plausible also makes it vulnerable to the conditional fallacy. A conditional fallacy involves ignoring the fact that, in a purported conditional analysis, the truth or falsity of the analysandum might depend on the truth or falsity of the antecedent of the conditional analysans (Shope, 1978). GI is open to counter-examples in which the conditional on the right-hand side of GI ("she would be motivated to ϕ") is true only when the antecedent of that conditional is false (i.e. it's not the case that A is deliberating correctly). These are counterexamples in which there is a reason to do something only because and so long as one is unsoundly deliberating or less than fully rational (Johnson, 1999).

An example will clarify. Johnson (1999) provides the following:

> On some ill-fated day you might come to believe that you are agent James Bond, licensed to kill, baccarat shark and connoisseur of fine port. And because of this belief you might come

[24] As a subjectivist view, GI endorses the view that the normative source of reasons is desire-based. But reasons justify actions insofar as they issue from the idealized agent, not her actual self. Objectivists will not like this. They will insist instead the normative source of reasons is value-based, and that the suitably idealized standpoint that generates normative reasons meets not only requirements of procedural rationality, but substantive requirements as well. I only want to point out this area of disagreement here so as to acknowledge the contentiousness of the claim in text. Chang (2014) is also useful to get a handle on the "problem of gridlock" between subjectivists and objectivists about normativity and normative reasons. See arguments in Chapters Two and Three for the motivation and justification of subjectivism.

> to have desires to play high-stakes baccarat, buy cases of rare wine and sneak around Russian embassies. Yet [GI] implies that there would be no reason for you to seek a psychiatrist, since were you fully rational you would not have this false belief and so would have none of these irrational desires. Far from having such thought disorders, you would not have any thought disorders at all; indeed, no lack of information, no bad habits or shortage of imagination. In sum, there could be no reason for you to do anything that seems appropriate for a person of diminished rationality. For were you fully rational, you would already be in the best rational condition you could be, and so could have no rational desire to become better or do things appropriate for a person of diminished rationality (61).

In this example, Johnson assumes that "James Bond" has a reason to see a psychiatrist to rid himself of false delusions about his identity. Since GI yields the result that "James Bond" does *not* have a reason to go to the psychiatrist, insofar as his fully rational self would not be motivated to do so (after all, he doesn't believe he is James Bond), Johnson concludes that GI must be wrong. Ultimately, Johnson challenges those who like GI to find some way around the conditional fallacy.

There are two obvious ways to avoid the conditional fallacy. The first is to eliminate the idealization process from conceptions of normative reasons. But this would amount to NI. It is not a defensible strategy. The second way to avoid the conditional fallacy is to move to a "wanting to want" theory of reasons, as proponents of IAA do, like those canvassed in the previous chapter. IAA predicts that James Bond has a reason to see the psychiatrist because his fully informed counterpart will want him to, even though James Bond would not be motivated to act on that basis. IAA is problematic, however, because it permits external reasons, reasons that are not capable of motivating or explaining the actions of actual individuals. Externalism allows individuals to be alienated from their normative reasons, as I argued in the previous chapter.

Johnson (1999) challenges internalists to defend a version of internalism that upholds the explanation constraint on reasons, but without giving up the normative constraint on reasons that renders it plausible in the first place.

5.3 Can Internalists avoid the Conditional Fallacy?

The motivating question is this section is whether internalists can meet Johnson's challenge. Can they avoid the conditional fallacy, but retain the explanation constraint on reasons, the hallmark of subjectivist-internalism? Indeed S-I's defensibility depends on this possibility. In this section, I will review three attempts to modify GI so that it avoids the conditional fallacy, only to show that they ultimately fail. I will propose an alternative response to the conditional fallacy.

5.3.1 Full vs. Practically Rational Selves

Michael Brady (2000) thinks he meets Johnson's challenge with the following proposal:

> There is a reason for A to ϕ in C only if (i) Af [fully rational A] would want A to ϕ in C, and (ii) Ap [practically rational A] would be motivated to ϕ in C if Ap were aware of the fact that (i), and aware of any other relevant facts (p. 96).

Brady believes this meets Johnson's challenge because it retains both the normative constraint on reasons and the explanation constraint, insofar as whatever A's reason is will serve to motivate *under the right conditions*, namely the condition of practical rationality (as opposed to full rationality). Brady says that formulations of internalism can forge a connection between what A has reason to do and what she would be (directly or indirectly) motivated to do if she were practically rational, rather than forge a connection between A's reasons and her present desires. On Brady's view, A's reason need not "make sense" from A's present perspective. Instead, Brady has it that the belief that Af would want A to ϕ in C would motivate Ap. Ap's belief "indirectly" explains A's action, even if A himself cannot be motivated by Ap's belief.

There are a few problems with this proposal. First, Johnson (2003) points out that it still commits the conditional fallacy: "'James Bond' surely has a reason to become aware of the fact

that his rational self desires him to stop believing he is James Bond. But he would not be motivated to become aware of it if he [‘JamesBond’p] were already aware of it” (p. 576). James Bond will still not be motivated to go to a psychiatrist because he is not practically rational, even if ‘JamesBond’p is motivated to do so. How much is built in to practical rationality determines whether practical rationality gives Ap reasons A lacks, and vice versa.

A further worry is that Brady’s proposal is not well suited to meet the explanation constraint. There remains a distance between Brady’s A and Ap – the current A and A when practically rational. Brady thinks that as long as Ap’s motivations could “indirectly” explain A’s actions, the explanation constraint will have been met. But it’s not clear that A could be motivated even “indirectly” to act upon the basis on which Ap is motivated. For example, ‘JamesBond’p may be motivated to go to a psychiatrist, but as long as James Bond isn’t even aware that he is not, in fact, James Bond, he will not be motivated to see a psychiatrist (unless, of course, he is persuaded that James Bond himself needs to go to the psychologist).

And finally, how does Ap become aware of Af? How does Ap become aware that his fully rational self wants Ap to ϕ? And why does Af’s advice matter to Ap? Brady does not provide an answer. To this end Johnson says, “it is no surprise that he offers nothing…Any condition in which ‘James Bond’p could respond to [Af’s desire] is not a condition in which he believes he is James Bond, and so is not a condition in which there is reason to remove the belief” (Johnson, 2003, p. 577).

5.3.2 Action Descriptions and Explanations

Joshua Gert (2002) proposes another way for S-I to avoid the conditional fallacy. He challenges readers to think about how the level of description of actions affects our intuitions about what someone has reason to do. Consider that someone sitting reading a book is doing

both (a) reading *Ivanhoe* and (b) doing something that gives him pleasure. Reading *Ivanhoe* is the token of doing something that gives him pleasure, which Gert calls a type of action. Now we ask the question: does Joe have a reason to do as he is doing? If reading *Ivanhoe* gives Joe pleasure, then having a reason to (b) is also a reason to do (a). A fully rational Joe may lack a specific reason to do (a), but still nonetheless have some motivation to do (b). If we suppose that (b) is a preferred action type of a fully rational Joe, then we can say that fully rational Joe is always to some extent motivated to do (b). What action token that indicates for current Joe is a matter of Joe's current preferences. Another way to understand the invocation of the token/type distinction here is by the fine-grainedness of the explanation of action. Action types refer to actions under a broad description (e.g. Joe is doing something that he enjoys); action tokens refer to actions under a narrower description (e.g. Joe is reading *Ivanhoe*). In sum, Gert's suggestion is that A's reasons will be tied to the type of action that Af [fully rational A] is motivated to do, and Af's motivations will figure in the correct explanation of A's action. His proposal says:

> There is a reason for A to ϕ in C [only if] there is some ψ (of the preferred type) such that A ψ's when A ϕ's in C, and a fully rational A would want to ψ (Gert, 2002, p. 93).

Gert names Af's preferred action types A's "basic ends" – those ends she would want to pursue if he were fully rational. He thinks that this iteration of internalism avoids the conditional fallacy because *gathering more information* (an action token) will fulfill Af's basic ends. As long as Af values as a basic end leading a happy life, for example, A will have reasons to gather more information, since that will advance A's progress toward successfully achieving happiness. If a person believes he is James Bond, he will have a reason to go to a psychiatrist because that is one token of a type of action preferred by Af, namely, leading a healthy, happy life, and living as James Bond thwarts progress toward that basic end. Note that there is nothing in Gert's proposal to suggest that all agents necessarily have or share certain basic ends.

This proposal, on first pass, does not commit the conditional fallacy, but it avoids it, again, at the expense of the explanation constraint on reasons. Gert does not tell us what conditions A would have to be in so as to "get" that her actions are a token of a type of action preferred by her fully rational self, and so motivated on that basis. James Bond, of course, does not believe that his belief that he is James Bond thwarts his living a happy, healthy life. What would have to happen to get James Bond to be motivated to give up his belief that he his James Bond? It would need to be more than reminding him that his fully rational self would like him to be happy and healthy. This alone will not motivate James Bond. Perhaps it is true that giving up on the belief that he is James Bond will cause James Bond to lead a happy, healthy life, something that the fully rational Bond wants. But delusional James Bond is utterly and completely incapable of believing that he should do what will promote his fully rational basic ends. The reason there is for James Bond to stop believing he is James Bond would not be James Bond's reason for doing so. So Gert's proposal fails to meet the explanation constraint on reasons, even though it may successfully avoid the conditional fallacy (Johnson 2003, 578).[25]

5.3.3 A Somewhat Less Idealized Account

Mark Van Roojen's (2000) proposal to meet Johnson's challenge involves "changing the specification in which motivation must be manifest so that it is less idealized…We can specify conditions which are ideal enough to ensure motivation but which are not so ideal as to be incompatible with the grounds of the agent's reasons" (p. 234). On first look, this strategy seems

[25] Perhaps in response to Johnson's (2003) criticism of his proposal, Gert seems to have backed away from itnernalism. Gert (2004) defends a hybrid approach to reasons and argues that we should be internalists about "requiring reasons" but externalists about "justifying reasons." I do not have the space to delve into the nuanced distinctions between requiring and justifying reasons, or how they map onto the conceptual landscape I mapped in Chapter One.

promising. After all, if the subjective motivational sets of not-so-ideal A (A*) and A overlap considerably, what motivates A* should also motivate A. Van Roojen's proposal is:

> An agent has a practical reason to do an action on certain grounds and in certain conditions only if, of the complete descriptions of that agent that include those grounds and circumstances, the ones that make the agent out to be most rational and relevantly informed include a motive to do that action.

This proposal does not require that A* is completely rational or completely informed. It holds fixed A's grounds and circumstances C – his actions to this point, his evidence, his beliefs, and his desires – and enhances his rationality only to an extent that is consistent with keeping C the same. The relevant grounds for which A acts are, if A is acting in accordance with reasons, the same as the grounds for which A* would also be motivated to act.

Johnson (2003) calls this a "piecemeal" approach (p. 579), "holding fixed the dimension in which we are irrational in circumstance – say, our false beliefs – and considering what we desire if we are *otherwise* ideally rational" (ibid.). In James Bond's case, Van Roojen's proposal holds fixed his false belief that he is James Bond and asks, if he were otherwise rational and relevantly informed, what would he be motivated to do? Van Roojen says, "holding fixed his actions to this point, his evidence, his beliefs, and his desires, it would be more rational to recognize his delusion and form the intention to see the psychiatrist than not to do so" (p. 238). The belief that he his James Bond is likely inconsistent with a whole bunch of other beliefs that the "most rational and relevantly informed" James Bond* endorses. So, Bond* will be motivated to give up the belief that he is James Bond, insofar as doing so makes the agent out to be most rational.

The problem with Van Roojen's approach is that it may devolve into NI. Johnson says:

> There could easily be an action ψ which would make A more rational in every way [e.g. gathering more information]…It seems that if there is a reason for A to do anything, there would be a reason for A to ψ. But van Roojen's view entails that there could be no reason for A to ψ. For were A fully rational in all of the ways that ψ-ing would promote, then he would not want to ψ. Yet if we hold fixed all of the grounds of this reason, we do not idealize…at all" (2003, p. 579).

Consider the example of a case where A falsely believes a glass contains gin. A relevantly informed A* already knows the contents of the glass, so there would be no reason for him to ask or check. Yet if we must hold A's circumstances – and his belief that it is gin – constant, then we do not idealize A at all. That is, A would not be motivated to check, since he thinks his belief that the glass holds gin is correct (Johnson 2003).[26] If we think that A nonetheless has a reason to check the contents of the glass, then we commit the conditional fallacy. If we deny that A has a reason to check the contents of the glass, we are, arguably, endorsing NI.

Van Roojen anticipates the worry that his "somewhat less idealized account" devolves into NI. In response, he reiterates that the normative constraint is retained on his account because agents could still be wrong on the account. It's possible that an agent will not gather information, reflect on it, and reason it light of that information despite there being an available route from his current subjective motivational set to a more rational and relevantly informed state. One may still fail to act on a motive to make himself more rational even if he would be motivated to do so under circumstanced he could be in. In other words, Van Roojen's "somewhat less idealized account" does not allow that "anything goes." I think this is a satisfying response, and a promising way to continue.

Despite providing a promising response, I think Van Roojen errs, however, in assuming that we are always out to be more rational. He thinks that we always have a reason to do

[26] As far as I can tell, Van Roojen has not responded directly to Johnson's objections to the "Somewhat Less Idealized Account" of reasons. Van Roojen continues to defend internalism about reasons, however, in Van Roojen (2005) on the basis that reasons must be capable of governing the choices of rational agents.

whatever action we are motivated to do also makes us out to be more rational. One may prioritize other goals, however. Van Roojen must provide an argument to support the statement that we prioritize becoming more rational rather than, say, doing what helps others, or satisfies our cares, or expresses our self. Van Roojen provides no such argument.

5.3.4 Do We Really Need to Worry about the Conditional Fallacy?

So far in this section, I've outlined and rejected three ways in which S-I could be formulated to avoid the conditional fallacy. Johnson (2003) ably showed in response to those attempts that evading the conditional fallacy requires tradeoffs – tradeoffs that S-I should be unwilling to accommodate. I think there is another way to respond to the conditional fallacy, though.

Let's return to the original test case: James Bond. Suitably idealized – or "fully rational" in the accounts reviewed – James Bond, James Bond+ is not motivated to go see a psychiatrist, since he does not believe he is James Bond. Yet, Johnson says, it's obvious that anyone who believes he is James Bond does have a reason to see a psychiatrist.

I worry about the "it's obvious" in the previous sentence. The assumption that anyone who believes that he is James Bond has a reason to see a psychiatrist invokes an externalist conception of reasons. We think that James Bond has a reason to see a psychiatrist even if he is not motivated to go. Perhaps more technically, we think this even though he would not be motivated to go strictly on the basis that he has psychotic thoughts. (One may be successful in convincing James Bond that James Bond has a reason to go to the psychiatrist, e.g. to get over a fear that interrupts his secret missions.) This assumption may jive with the common hope that

psychotic individuals get much-needed help, but that establish the truth of the intuition that he the same people – or this particular person – has a reason to do so.

The conditional fallacy is a problem only when we think that a "suitably idealized" James Bond will no longer have delusions about his identity. This is predicated on the idea that suitably idealized James Bond has access to, and will believe, the fact that he has delusions in the first place. But why think that James Bond will accept that? We think that the suitably idealized James Bond will agree with us, and Brady, Gert, and Van Roojen have built conditions to make sure that he does. But, unless there is some independent motivation for idealizing James Bond *in that way*, we could spell out a different "idealization process" entirely, one that ensures that our suitably idealized selves would be motivated to act in ways that substantially overlap with our non-ideal selves, in ways that our non-ideal selves could act. In other words, we should consider, and allow, the real possibility that James Bond+ still would not be motivated to go to a psychiatrist because, even under suitably idealized circumstances, he persists in the belief that he is James Bond. The job of a theory of reasons is not to settle what a particular person does and does not have reason to do. It is instead to defend the truth conditions on reasons-statements. We should not presuppose an answer to the question of what someone can and cannot have reasons to do in that pursuit.

5.4 Reasons as "Normativized Explanations"

My strategy going forward is to show that the even "suitably idealized" James Bond will share enough of James Bond's subjective motivational set that what James Bond+ would be motivated to do will be "on the table" so-to-speak for James Bond, too. We shouldn't expect or assume that the correct or best "idealization process" will ensure that James Bond+ is motivated to go to a psychiatrist. The claim that our suitably idealized self's subjective motivational set will

overlap substantially with the non-ideal self's set secures the explanation constraint upon reasons. The explanation constraint says that what A's reasons are fixed by what A would be motivated to do under suitably idealized circumstances, and also that A herself could act on the basis of her reason. In sum, it says that the reason there is for A to ϕ could also be *A's* reason for ϕ-ing. Setting up the "idealization process" in a way that respects the explanation constraint entails that the "suitably idealized" A, A+, and A will share enough of the same subjective motivational set and beliefs that whatever A+ would motivated to do, A could be brought to motivated to do as well. But how might we defend the explanation constraint?

Defending the explanation constraint is a crucial argumentative move for the internalist. Sobel (2001a) accused Bernard Williams of advancing the thesis of internalism as its only defense. He says, "Williams's explanation condition… which he took to be the centerpiece of his case for internalism turns out to be...the thesis of internalism" (223). Sobel's charge is that internalists often advance the explanation condition as a desideratum of a theory of reasons, and propose a definition of reasons which captures that desideratum, but without sufficiently motivating or defending this desideratum in the first place. Why keep it, especially given the challenge of avoiding conditional fallacy?

In the remainder of this chapter, I will critique Williams' defense of the explanation constraint. Williams draws from Davidson's thesis (1963) that reasons are, fundamentally, explanations, and adds that normative reasons are best understood as "normativized explanations" (Finlay, 2009).

5.4.1 Reasons as Explanations

Davidson (1963) advances thesis that reasons, if they are anything, are explanations of an agent's action. Adding explanation shows not merely what from another person's perspective

could count in favor of acting, but why that person did, in fact, act. Davidson (1963) says that in giving reasons, we point to "some feature, consequence, or aspect of the action the agent wanted, desired, held dear, thought dutiful, beneficial, obligatory, or agreeable" (p. 685). And so, while the purpose of reasons is to provide evidence for why you should do something, simply stating that the evidence "counts in favor" does not give us any indication that the beneficial or desired feature is a reason for you. Even if we could agree on what counts in favor of, say, hopping on the next flight to Chicago, we need not agree on what among that class of considerations are reasons. There are lots of considerations that might explain why one would go to Chicago: her mom lives there, and needs assistance, she was invited to a wedding there, she's interested in studying skyscrapers, or is a member of the Baha'i faith, or she's never seen the tiny rooms exhibit at the Art Institute. There are an infinite – or at least a very, very large finite – number of considerations that count in favor of hopping on the next flight to Chicago. Davidson says that what characterizes reasons specifically is the ability of any of these considerations to explain someone's behavior.

We must be careful here to distinguish here between kinds of explanations. I can explain my going to the store to get milk by recounting the succession of biological and physical events that occur. This is not the kind of explanation that Davidson has in mind as a defining feature of reasons; this is not an appropriate answer to the question, "why did you go to the store?" The appropriate response, rather, explains genuinely why *you* would go to the store: you ran out of milk. This is your reason, and it at once counts in favor of going and explains *your* action. The relation between agents and reasons is emphasized through explanation. Reasons do not explain why anyone would go to the store, but rather why you or I would go to the store.

Davidson builds into his account that considerations that explain the action or "rationalize" the action in the relevant way also cause the action, which has given rise to a rich debate whether reasons for action are psychological states that cause action (i.e. beliefs and desires) or an external fact or consideration (i.e. the content of a belief).[27] Thankfully, Hieronymi (2011) convincingly clarifies and reinterprets Davidson's causal claim. Hieronymi argues that we should interpret Davidson as saying that when we explain an action in such a way as to make it clear that the action was done for a particular reason (e.g. "we ran out of milk"), we do so by noting that the agent, for that reason, settled for himself or herself the question of whether so to act, therein intended so to act, and executed that intention in an action that was the event in question. We need not interpret Davidson to be saying that a reason refers to a physical, psychological or mental state that (literally, or metaphysically) causes an action, but rather that a reason picks out certain elements of subjective motivational set, or considerations that figure in explaining why *this agent* so acts. Interpreted this way, reasons pick out the considerations for which one acts; reasons serve as motives and as explanations.

5.4.2 "Normativized Explanations"

To be sure, Davidson's account of reasons as explanations is not an account of *normative* reasons; it tells us only that practical reasons, properly understood, explain our actions. According to one interpretation of Williams, Williams sought to extend Davidson's remarks about explanatory reasons to normative reasons. According to Williams, to believe that R is a (normative) reason for action *is nothing other* than to believe it is a special kind of explanation

[27] See especially Anscombe (1983) and Dancy (2000).

for acting (Finlay 2009).[28] A reason is an explanation not of why an action actually was performed but of why it would be performed under certain conditions. I've been calling these conditions those of "suitable idealization" or "suitably idealized" A so as not to assume that one idealization process is correct. Williams terms his "idealization process" the conditions of sound deliberation.

On this account, when an agent acts for normative reasons, those are the reasons for which she acts, and they serve to explain the action. Reasons, then, are considerations that figure in explaining an agent's action. They also serve, though, to justify the action, and so they remain *normative* reasons. In this way, Williams' proposal follows Davidson's (1963) suggestion that reasons must explain an agent's acting. Reasons for Williams offer a special kind of "normativized" explanation. Of course agents often act for "reasons" which are not normative; they explain action but rest on false beliefs, for example. So normative reasons are not *simply* explanations, but explanations under a certain set of conditions – those of sound deliberation. The notion of a normative reason is a consideration that would explain an agent's action under the condition of sound deliberation. Finlay explains:

> To say that *this glass contains gin* is a reason for A to drink for it, for example, is to say that it explains why A would drink from it under the condition of sound deliberation. The first premise of Williams' argument is therefore that 'R is a reason for A to ϕ' means that *R is an explanation of why A would be motivated to ϕ if he deliberated soundly* (Finlay, 2009, p. 14).

So far, so good. Putting aside for now exactly how Williams specifies the conditions of sound deliberation, I'm going to continue with his argument for the explanation constraint. He must first establish that the explanation constraint is desirable and defensible.

[28] In this section, I rely heavily on Finlay's (2009) interpretation and explication of Williams' account and defense of internal reasons. I do so because he convincingly argues that it is the most charitable and reasonable interpretation of Williams' account. See Finlay 2009.

The ability to explain action, Williams thinks, sets reasons apart from other kinds of normative statements, e.g. that donating money to charity is the right thing to do. If in telling A that she has a reason to give to money, we tell her that there is an explanation why *she* would give or of how giving is connected with some element of her subjective motivational set, A is more likely to respond. She will see that giving money satisfies not just some abstract moral obligation but also achieves something she values (perhaps only fame). On this point Williams says, "what we can correctly ascribe to [A] in a third-personal internal reasons statement is also what [A] can ascribe to himself as a result of deliberation" (1981, p. 103), and "internal reasons statements can be discovered in deliberative reasoning" (1981, p. 104). A reason for A to ϕ is therefore an explanation *for A* of why she would ϕ, i.e. something that could be explanatory to A were she deliberating correctly about how to satisfy an interest she already has. Here Williams draws upon the now-familiar argument that a statement of reasons for A must be distinctively about the actual A and responsive to *her* concerns and interests. The conditions of sound deliberation that he adopts are responsive to her concerns and interests, whatever they are, not the interests of abstract others or an impartial perspective.

To defend the explanation constraint, Williams then asks whether external reasons can be an explanation *for* A of why she would ϕ. Recall that external reasons, if they exist, do not and need not motivate or explain an agent's actions, and so the statement that I have an external reason amounts to accepting: "I have a reason to ϕ*,* even if it would not explain my action or make it intelligible to me." Williams asks: could we accept that statement if we know absolutely nothing about ourselves or the contents of our subjective motivational sets?

According to Finlay (2009), Williams' case against external reasons hinges on what A comes to believe when she believes she has a reason to ϕ. When it comes to playing golf, for example, A may come to believe one of three beliefs:

1. The non-normative belief that R, e.g. that golf is played outdoors.
2. The specific normative belief that R is a reason for A to play golf.
3. The general normative belief that R is a reason for A to play golf.

Believing merely *that golf is played outdoors* may count in favor of playing golf, but it is not itself a reason in every instance. So when A comes to believe that it is a reason for playing golf, she believes more than the contents of (1). (3), though, says too much. A "general normative belief" that golf ought to be played because it is played outdoors amounts to asserting the value of outdoor sports. But even endorsing that outdoor sports are valuable in general does not establish that they are valuable for A, nor that its value serves as the basis for why A acts. So (3) cannot be what A believes when she comes to believe that *she* has a reason to play golf. But when (2) is satisfied, A believes that R is a consideration could explain her action; it connects to the action to something in her subjective motivational set, namely a desire to play sports outdoors. R is, therefore, an internal reason for A, namely when A believes that R is a consideration that would explain why she plays golf.

Williams point is that it is a mistake to think that the possibility of external reasons is proven merely by pointing out that A is motivated by the fact that golf is played outdoors. If external reasons exist, its proponents must show further that the agent acts on the basis of the belief that R is a reason for which he acts, and that R genuinely explains his action, absent information about himself; and that A is motivated to act on R simply because he is deliberating from the impartial perspective, that some algorithm determined it was best, or a citizen in the kingdom of ends would be so motivated. Which is to say, with no additional information about

who *he* is and what *he* wants, independently of anything in his present motivational set, of why he would be motivated to ϕ. But, says Williams, this is impossible, because "there is no motivation to deliberate from to reach this new motivation" (1981, p. 109). Finlay reinterprets this claim to mean:

> There are no *facts* about the agent's motivational set that warrant his inferring the *fact* that he has an (external) reason, i.e. that there is an explanation why he would be motivated to ϕ by deliberating soundly whatever his motivational set...If an agent cannot discern in his psychology even a disposition to be motivated to ϕ by R, then he cannot properly conclude that R is an explanation for why he would ϕ if he deliberated soundly, and hence a reason for him to ϕ" (Finlay, 2009, p. 18).

We only come to believe that something is a reason when it is a reason for me when I believe that R could genuinely explain why I would act. But in order to believe that, much less justify it, one must know something about oneself. I couldn't believe truly that something could explain my actions if I didn't consider its relation to my desires. If I scanned my motivational set, I would find no evidence that "golf is outdoors" would explain why I play it. (I happen to play golf because it affords me uninterrupted time to hang out with my dad.)

So, while it might be true that there are facts or proposition that would motivate me if I found about them, and so in some sense it seems that there are external reasons, once I come to believe that fact is a reason, it ceases to be external, it becomes a reason *for me*, an explanation of why I would act. It would now explain why I am motivated to A. But, that I come to see it as a reason, that it is true that I have a reason to A, is true only in virtue of a psychological link between states of affairs and me. There could be no other way of defending the truth of the statement; "I have a reason to ϕ," except by citing how ϕ-ing is related to my own desires, concerns, motivations, etc. Williams' account amounts to an error theory of external reasons; it is an empty category. What the term 'reason' refers to are reasons *for me*, or internal reasons. External "reasons" may be any consideration, norm, moral principles, etc. proffered to steer my

action. But unless those considerations, norms, principles, or facts would also genuinely explain my action (under conditions of sound deliberation), they are not reasons.

5.4.3 Sound Deliberation

Having provided Williams' argument for rejecting the existence of external reasons, I will go on to state what he means by "sound deliberation." Reasons are "normativized explanations" that would explain the behavior of agents if they were soundly deliberating. What does that require?

Williams allows that sound deliberation from one's subjective motivational set about what would explain one's action manifests in various ways. He offers few guardrails on what counts as deliberation. Suffice it to say, though, that sound deliberative practices go beyond instrumental reasoning. He writes, "the mere discovery that some course of action is the causal means to an end is not itself a piece of practical reasoning," whereas "a clear example of practical reasoning is that leading to the conclusion that one has a reason to ϕ because ϕ-ing would be the most convenient, economical, pleasant, etc., way of satisfying some element in [her motivational set]" (1981, p. 103). Instrumentally reasoning from my desires to the courses of action that satisfy them is not enough to settle what I have reason to do. Williams provides two further conditions for sound deliberation: (1) improved information and (2) imagination. Improved information is like the requirement for full information, but with an important caveat. An agent's epistemic circumstances and subjective motivational affect what she can come to know, and so are relevant to what the agent has reason to do. Williams' inclusion of imagination as a part of sound deliberation underscores his commitment to pluralism about what soundly deliberating looks like in practice.

5.3.4.1 Improved Information

When we deliberate, or reason practically, Williams says we need at least true information that is relevant to the circumstances at hand. We must be factually and correctly informed in order to deliberate soundly about what we ought to do. Deliberating soundly requires correcting misinformation that bears on what my reasons are. Williams' aforementioned gin-and-tonic case is the paradigm example of this. Since there is no gin in the glass, there is no reason for the gin-seeker to reach for it. There is no sound deliberative route from his subjective motivational set, which contains the desire to drink gin, to drinking this glass of petrol, since he has no desire or concern for drinking petrol. If he corrected his misinformation, he would realize he has no reason to drink the class in front of him. Thus, soundly deliberating about what we ought to do requires that we have true information that is relevant to successfully satisfying our desires. One might ask: why analyze reasons in terms of these conditions, namely the absence of false belief, rather than, say, the possession of moral desires or conversion to an impartial perspective? Among her interests, Williams says, is being "factually and rationally correctly informed" (1995a, p. 37) Williams thinks that this is an interest that all agents share.

Williams has been criticized sharply on this point on the basis that he cannot assume that all agents have this desire (Sobel, 2001a). The charge is that if Williams is unwilling to grant, for example, that rationality requires that agents value humanity, he is not entitled to assume that agents have an interest in being factually and correctly informed. One reason for thinking this condition is justified, however, is by the very nature of desires and other elements of our subjective motivational sets (A.H. Goldman, 2009, p. 60). To desire is to want to satisfy desire. But desire can be satisfied successfully only if we know how to do so. Having a desire for anything thus entails a desire for true knowledge about one's surroundings, in particular elements

of one's surroundings relevant to satisfying desire or meeting concerns. Successfully satisfying desires requires correct mind-to-world fit, and so desiring entails an interest in correct mind-to-world fit.

Despite agreeing that factual and correct information is a condition of sound deliberation, Mason (2006) argues that Williams also acknowledges that there may be a limit to what we can expect agents to know or grasp about a given situation. Williams specifically mentions "unconscious obstacles" or "blocks" (1995b, p.188) that matter when determining what a person has reason to do.

Williams, unfortunately, does not expand much on the implications of these "blocks" for his theory of reasons. Mason (2006) picks up this task. Williams does not say whether that these blocks make it impossible for agents to soundly deliberate or whether, instead, they render reasons relative to agent's epistemic and social circumstances. Both ways of interpreting his claim are available. Let's consider these two options.

It could be said that these "blocks" make it impossible for agents to meet the conditions of sound deliberation. Agents whose religious beliefs and motivations render them unable or unmotivated to accept certain scientific facts, for example, will resist accepting scientific information relevant to their decision. This person, it might be said, will fail every time she deliberates over what she has reason to do when religious beliefs indicate something inconsistent with science in virtue of the fact that she is in violation of Williams' truth norm for sound deliberation. This will have the result of alienating agents from their reasons, from what they would be motivated to do if they were soundly deliberating, since they cannot soundly deliberate.

A relevant example from bioethics literature concerns Christian Scientists. Christian Scientists deny that modern medicine is effective in treating illness. Motivating this opinion is

the belief that illness results from "human alienation from God" produced by a fundamental misunderstanding of one's spiritual nature. With this understanding of illness, they maintain that prayer through Christian Science healing is the best medicine. The belief about illness, combined with the strong inclination to be a Christian Scientist presents a "block" that would have be to be removed in order for a Christian Scientist to arrive at, say, the motivation to take antibiotics. It's important to note that Christian Scientists do not believe that modern medicine is immoral, but deny that it is efficacious in treating illness. This makes their thoughts about medicine different than say, a Jehovah's Witness, who believes that blood transfusions do work, but also render the recipient unclean or impure (Battin, 1999). In this way, the Christian Scientist's beliefs fly in the face of modern medicine. In deliberating about what to do about a sinus infection, that amoxicillin is an efficacious treatment, simply will not enter into the deliberative process of a Christian Scientist. The "block" of religious conviction would have to be removed in order for the Christian Scientist to come to the conclusion that she has a reason to take amoxicillin, even though every doctor would insist that she does. Does this block render her incapable of deliberating soundly?

I think not. As Mason (2006) argues, what an agent has normative reason to do is limited to what motivations it is possible for the agent to arrive at, blocks and all. It is not possible for the Christian Scientist, given the "block" her religious convictions present, to accept the tenets of modern medicine. And so, it is not possible for those tenets to enter into her deliberative processes. The information upon which she bases her decision, e.g. beliefs about the etiology of illness, while seemingly false, do not qualify as misinformation that needs to be improved or corrected from the perspective of the Christian Scientist. Furthermore, she is still capable of imagining what it would be like to be healed by God, or conversely, to not be healed, and still

capable reflecting on what's important to her. The relevant information used when soundly deliberating is constrained by her "block," but that does undermine the normativity of her reason for visiting a Christian Science healer (rather than take amoxicillin).

An agent needs information relevant to making her decision in order to successfully deliberate. We cannot expect certain agents, however, to accept information that would require giving up elements that are central to their subjective motivational sets or information that is impossible for that agent to acquire given epistemic and social circumstances. I return to the gin-and-petrol case, slightly adapted in ways suggested by Mason (2006, p. 169). Let's call the gin-seeker Drew. Whether or not Drew has a normative reason to drink the contents of the glass in front of him depends on whether he would be motivated to do so if he were soundly deliberation. In turn, in order to know whether Drew has a reason to drink the glass, we must ask: what information Does drew need to soundly deliberate? Is that information that Drew can access? Or does Drew have some "block" that renders him unable to update the belief that it contains gin-and-tonic? Though perhaps unsatisfying, my answer is that it depends. If Drew further believes that his loving and honest partner put the glass there, then it may be unfathomable that the glass should contain anything other than gin-and-tonic. In this case, I maintain that it's possible that Drew has a normative reason to drink the glass, insofar as he could not revise the belief that it contains gin without giving up a strong conviction, namely, that his friend has his best interest at heart. Sound deliberation would not require updating information about the contents of the glass, since doing so would require giving up a deeply held conviction about his partner. If, on the other hand, Drew's mischievous friend placed the glass in front of him with a laugh, a soundly deliberating Drew would have a reason to at least test the contents of the glass.

Sound deliberation requires that we update information that bears on what reasons we have only when doing is not inconsistent with other convictions or deeply held beliefs. But we can do this only up to a point – the point at which accepting certain information requires alienating oneself from a deeply held conviction or value. In other words, we are sometimes "blocked" from accessing and appreciating even true information because it conflicts with elements of our subjective motivational set. The requirements of sound deliberation are responsive to who we are and what we care about.

5.4.3.2 Imagination

Williams allows that sound deliberation can take other more amorphous forms. He explicitly includes imagination as a form of practical reasoning. In order to assess whether we have a reason to perform a particular act, we should think about what it would be like to act on that reason. Through deliberation or conversation we may come to have "some concrete sense" of what would be involved in satisfying some element of one's motivational set or acting on the basis of a particular reason (1981, p. 105). Williams continues that in imagining an agent may, "lose his desire for [some outcome], just as, positively, the imagination can create new possibilities and new desires" (ibid.).

Suppose one is considering whether to attend a concert or whether to go out to eat with a group of friends. In deciding, it would help to imagine what it would be like to be at the concert. Suppose the last concert you went to with your favorite band was over-crowded and rowdy. This experience, which informs how the next concert will be, might render you less inclined to go, especially, for example, after a particularly stressful week. In the mood for a more relaxing evening, imagining sipping wine over carefree conversation strengthens the inclination to go out

to dinner. It's now apparent that catching up with old friends over wine and conversation is, all things considered, what you ought to do tonight. Imagination was a crucial part of the process of deliberating over what's the best thing to do, and this deliberative process resulted in the loss of the desire to go to the concert and the strengthening of the desire to go to dinner instead.

Thinking about states of affairs, or imagining what it would be like to be in certain states of affairs, allows us to prioritize certain desires, concerns, or preferences or rid ourselves of others. It not only acts as a way to instrumentally reason from ends to means, but as a way of finessing or honing the contents of our motivational sets. As a friendly amendment, Smith (2012) suggests that we make a distinction between motivational and affective components of desire in order to make sense of the role of imagination in Williams' account. The affective component of a desire is "lik[ing] it's being the case" (Smith, 2012, p. 213). In other words, desiring means we are affectively oriented toward the world's being a certain way. We may like it being the case that our social events are subdued opportunities for talking, or we make it like that our social events are rowdy affairs. Or, we may be affectively oriented toward both of these states of affairs, just at different times or under different circumstances. We have this affective desire or disposition regardless of whether we are planning social events. The motivational component of desire is "the disposition to making something the case in circumstances in which you have the opportunity to make it the case" (ibid., p. 212). Motivational dispositions compel us to perform actions that make it the case that our affective desires are satisfied, when we are capable of doing so. For example, if you like it to be the case that your social events are subdued affairs that facilitate talking with friends, that your friend is organizing a birthday dinner will incite the motivational disposition to attend, and it will at once explain why you would go to the birthday party. That the concert will be rowdy dampens the motivational disposition to attend, and that the

concert is happening tonight will no longer be a reason to go. Smith characterizes this process when he says:

> Agents who deliberate correctly will subtract a motivational disposition from their [motivational sets] when they discover, via an exercise of the imagination, that they wouldn't like it if the world were the way they are disposed to make it, and they will add a motivational disposition to their [motivational sets] when they discover that they would like it if the world were a way that they can make it (ibid., p. 215).

When it is pointed out through imaginative exercises, discussion with others, or improved information that certain states of affairs, e.g. the rowdy concert, do not satisfy affective desires, then there should no longer exist a motivational disposition to pursue those states of affairs.

This imaginative process sets Williams' notion of deliberation and idealization apart from instrumental means-end reasoning on the basis of true information or the "desire-in/desire-out" model of ideal advisor accounts of reasons. Williams indeed leaves room for deliberating agents to acquire new desires or change old ones. As we imagine, weigh considerations, talk with others, learn information or remember previous experiences that are relevant to our decisions about what to do, our desires change. Though what someone has reason to do is limited by her antecedent desires, including affective and motivational dispositions, it is not limited to what she is motivated to do at the start of deliberation. Indeed agents can and do acquire new desires and hone old ones in the process of deliberation. Williams' account of deliberation allows for ways of "finessing" the desires and concerns and attitudes that generate those reasons. We have reason to do what we would be motivated to do upon completion of that process. But, crucially, what the process involves is still sensitive to "blocks" to one's imagination and epistemic circumstances.

5.5 Manne's Modifications

Before addressing objections to Williams' argument, I want to register that Kate Manne (2014) essentially endorses and expands upon Williams defense of internal reasons in her aptly named paper, "Internalism about Reasons: Sad but True?" Manne's proposal respects the explanation constraint upon reasons. The main difference between Manne and Williams is that while Williams says that reasons are generated by the desires of a suitably idealized agent, Manne says that the standpoint from which reasons are fixed is not the agent's own suitably revised perspective but the vantage of a second person. Manne owes the inspiration for this view to P.F. Strawson (1962).

According to Manne, a reason is a consideration that an ideally situated advisor would want you to be moved by. Despite using the same term ("ideal advisor") as Ideal Advisor Accounts of reasons, the ideal advisor of Manne's account differs dramatically. Manne's ideal advisor is:

> Imagined to be a flesh and blood human being – as opposed to the disembodied voice of reason – who is ideally suited to play this social role, partly in being possessed of all the relevant information and being fully procedurally rational (or at least as fully procedurally rational as any actual human being could be). She might also be imagined to be virtuous and wise, perhaps – and she must at least be well-disposed toward her advisee. Finally, she should be imagined to be especially well-suited to play this social role for the particular agent in question. She is the person who is best suited to 'getting through' to her morally...The role of the ideal advisor is to persuade or recommend, not to issue *de facto* commands to the deliberating agent (p. 97).

In insisting that A's ideal advisor is another person, a flesh and blood human being, I interpret Manne as saying that sound deliberation on its own is not enough to establish the truth about what we ought to do. "Reasoning with" another person is a better way of reasoning practically about what one ought to do. On Manne's view, reasons for an agent A to ϕ are considerations which A's ideal advisor would be apt to cite in favor of A's ϕ-ing. She says, further,

> A consideration would only be apt to be cited in favor of A's ϕ-ing in this context if A would be (somewhat) motivated to ϕ, by the end of the conversation. It follows that the reason for an agent A to ϕ can hold only A would be (again, somewhat) motivated to ϕ, at the end of an idealized process of being reasoned with in this way by her ideal advisor. She must have the relevant motivational propensity, is a convenient way of putting it (p. 112).

In this way, Manne endorses and retains Williams' internalism. What the ideal advisor recommends to her advisee is constrained by motivations that the agent would have at the end of the conversation with her ideally situated advisor. More abstractly, the process of reasoning with someone will accordingly be constrained by those motivations that the agent herself would have if the process of reasoning with her were to be perfected and completed. And so, what a reason is constrained by the motivations of the agent herself. Reasons are, after all, considerations that one's ideal advisor would want her advisee to be motivated by. Things that would not motivate her after deliberating and reasoning with her are off the table.

While I think that this move does not belie subjectivism or internalism, it is odd for a subjectivist theory to prioritize the perspective of a different subject than to whom those reasons apply. The most compelling argument for internalism is that it avoids alienating people from their reasons; reasons must make sense to the person to whom they attach. Fixing reasons to a suitably idealized, soundly deliberating first person perspective seems to capture that subjectivist insight better than Manne's second personal account.

The virtue of Manne's account is the recognition that reasoning is often an interpersonal activity. With the expansion of "sound deliberation" beyond mere instrumental reasoning, Williams opened the door the idea that novels, movies, imagination, trauma, stimulating discussions go a long way toward finessing our desires and pointing out not only what we ought to do but who we want to be. Manne's proposal captures that that process invariably involves other people, notably those who are older and wiser, we might say, but also those we trust are

caring and looking out for us. Talking with them, discussing with them, reasoning with them, and the advice they proffer in practice often yields clarity about what we ought to do. To insist that the second person's determination is the only way to establish what one ought to do is exclusionary of the other types of deliberative activities that fall under soundly deliberating. How we arrive at decisions, new motivations, and the conclusion that we have reason to ϕ rather than ψ is often (perhaps even usually, or for some, always) the result of interpersonal discussion and influence. But it need not always be the case. For this reason, I think a view privileges the first person perspective, without denigrating the importance of interpersonal interactions in developing that perspective, is preferable. [29]

5.6 Limitations: The Need for a Stronger Defense of the Explanation Constraint

Those who disagree with Williams and Manne will do so on the basis that an agent can correctly recognize – and believe – that a consideration is reason, even for herself, absent personal information about her subjective motivational set. Parfit (2011) says that we simply must accept that we have certain reasons for action, whatever else we may want. McDowell (1995) says that in deliberating about what reasons we have, we come to "consider the matter aright" (p. 73). Once we realize what a correct deliberator would do in our situation, we ought to adopt *that* as our normative reason for acting, whatever else we may want. Or, as we saw, in a neo-Kantian tradition Korsgaard (1986) advocates a view whereby deliberating correctly requires adhering to certain principles, like the Formula of Humanity, that indicate what we would be motivated to do if were reasoning correctly. What these approaches have in common is that

[29] In addition to this objection to Manne's account of reasons, I raise an objection to her argument for S-I in the next chapter. Manne's (2014) argument for S-I depends on the truth of motivational internalism. Motivational internalism says that recognition of one's reasons is intrinsically motivating. This premise is sometimes used as a justification for reasons internalism, the idea that the truth conditions of something being is a reason is its ability to motivate action. Hampton (1998) persuasively argues that one cannot use motivational internalism as a justification for reasons internalism (see esp. p. 58). I agree, and will expand on this in Chapter Six.

sound deliberation does not involve deliberating about our actual desires or motivations, but deliberating about what we *would* desire or be motivated to do under specified ideal or rational conditions, unencumbered by our practical identities and idiosyncratic desires. (They differ, of course, in spelling out the ideal conditions.)

What we *would* do under better conditions, though, cannot explain what we in fact do in the present, since there are significant psychological differences between the properly brought up person, the citizen of the kingdom of ends, and me. These differences matter in considering what would explain my own acting, as opposed to explaining a different person's acting. That I'm not deliberating under ideal conditions matters. Williams said in response to McDowell, "I know that if I fall short of temperance and am unreliable with respect to some kinds of self-control, I shall have good reason not to do some things that a temperate person could properly and safely do" (1995b, p. 190). This response reiterates that what I value, desire, or am concerned about affects what my reasons are. And since reasons are "normativized explanations" and, absent information about one's motivational set, one could not arrive at the conclusion that an external consideration, fact, or principles would genuinely explain his action. Sure, it might explain why someone citizen of the kingdom of ends acts in a certain way. But that's not *me*.

Finlay (2009) raises a related concern that the conceptual tie between reasons and explanations is unwarranted. He says, "Even granting that the concept of a reason just is the concept of an explanation, it doesn't follow that a normative reason for action is any kind of explanation of an agent's acting" (p. 20). He says it may instead be an explanation of why it would be good to act, or why a better person would act. It doesn't follow that what I have reason to do explains why *I* act. Though Williams sketches (however sketchily) an answer this challenge, Finlay is obviously unconvinced. To salvage the conceptual tie between reasons,

explanations, and the first person perspective, I offer a novel argument for the explanation constraint on reasons in the next chapter.

Chapter 6: Expressive Reasons

I ended Chapter Five by noting objections to Williams' defense of the explanation constraint. At worst, Williams is accused of assuming the truth of the explanation constraint (Sobel, 2001a), at best, he is accused of misconstruing what it requires (McDowell, 1995; Finlay, 2009). Objectors ask: even if the explanation constraint is a defensible condition on reasons, why think that reasons must be "normativized explanations" of actual persons' actions rather than, say, explanations of what would be good to do or what one would do if she were rational or brought up differently? Williams' critics do not grant that he has provided sufficient support for the explanation constraint upon reasons, the hallmark of a subjectivist-internalist (S-I) view.

In this chapter, I provide another argument for the explanation constraint. Proponents of S-I need a better argument for the explanation constraint on its own. Arguments put forward by Williams and Manne have not convinced fellow subjectivists, much less objectivists. Proponents of S-I must also do something else, too. S-I theories of reasons should be able to tell us which desires generate reasons, and provide a principled reason for thinking that certain desires or motivations or other elements of one's subjective motivational set provide reasons but not others. Does any fleeting whim that survives the idealization process generate reasons? If I have a sudden impulse to grab the candy bar in the checkout line did I have a reason to do that? Why or why not? S-I should be able to provide an answer, and explain why someone has (or doesn't have) a reason to grab the candy bar. So, in rebuilding an S-I account of reasons, we must defend

the explanation constraint, and defend a principled reason for saying that a particular subclass of desires, motivations, whims, preferences, and other elements of one's subjective motivational set generate reasons but others do not.

Finally, we must do these two things without begging the question in favor of either subjectivism or internalism (a tall order, indeed). We need a better argument for the explanation constraint because, some argue, it is best understood as saying that reasons explain, rather, what would be good to do rather than explain why I would act. This brings us back to the problem of gridlock. Some remain firmly planted in the camp that says that reasons are grounded in objectivist moral considerations or concerns, while others say that reasons are grounded in desires. This chapter turns away from asking the very contentious question, "What is the source of the normativity of normative reasons?" and asks instead, "What do we do when we act?" Reasons for action should, after all, have some connection to action. I want to explore what mileage we can get out of changing the starting point of inquiry into normative reasons. So, rather than starting out with a question about reasons per se, I start out with an analysis of action. This move is similar to a move seen in Chapter Three made by Velleman (1996) and A.H. Goldman (2009), but I provide a different account of action, eschewing the notion that action has a particular constitutive aim.

Starting with observations about action, I argue, we arrive at a stronger argument for the explanation constraint, and a neat way of determining which desires generate reasons and which do not generate reasons. And, I hope, starting with an analysis of action does not beg the question in favor of either subjectivism or internalism. I argue that the connection between self-expression and action secures the explanation constraint on reasons, and so secures S-I about reasons.

My distinctive account of reasons "Expressive Reasons" (ER) says that R is a reason for A to ϕ when ϕ-ing is an expression of soundly deliberating A's self; and A, under ordinary conditions, would act on the basis of R. In closing, I provide responses to objections to ER. I countenance the "anything goes" problem by sketching an alternative account of the relationship of normative reasons to moral theory. The comments at the end of this chapter are admittedly speculative, and deserve much greater attention. My hope is that my defense of subjectivist-internalism and the strength of the arguments for ER establish a secure foundation for exploring its implications in future work.

6.1 Self-Expression & Action: An Argument for the Explanation Constraint on Reasons

Philosophical analysis of expression has occurred in silos, confined to aesthetics (Collingwood, 1938; Croce, 1901/1992; Wollhiem, 1993) or philosophy of language (Ogden and Richards, 1923; Green, 2007; Davis, 2003) or ethics. Within ethics, expressivists argue that terms like 'good' and 'bad' express attitudes, rather than ascribe properties (Gibbard, 1990; Blackburn, 2006). Some have suggested a connection between self-expression and moral responsibility (Sripada, 2015). And, there is some attention to emotional expression and actions that express emotions (Betzler, 2007). Surprisingly little attention has been paid, however, to the wider set of actions that expressing one's subjective motivational set that includes, among other elements, desires, concerns, pro-attitudes, cares, evaluative states, and other moods. This is surprising since no one could doubt that action is one of the most important and common ways that human beings express themselves. Elizabeth Anderson's (1993) expressive theory of rational action is one exception. I will discuss her view in what follows.

I will argue that insofar as actions express elements of one's subjective motivational set, there is an inviolable tie between actions, reasons, and the first-person perspective. This connection places a constraint on the kinds of considerations that can serve as normative reasons. The argument ultimately provides a defense of the explanation constraint on reasons.

6.1.1 Self-Expression

"Expression" is a difficult concept to define. Indeed, Ogden and Richards remarked over ninety years ago,

> It is certainly true that preoccupation with 'expression' as the chief function of language has been disastrous. But it is not so much because of the neglect of the listener thereby induced as because of the curiously narcotic effect of the word 'expression' itself. There are certain terms in scientific discussion that seem to make any advance impossible. They stupefy and bewilder, and yet in a way satisfy, the inquiring mind, and though the despair of those who like to know what they have said, are the delight of all whose main concern with words is the avoidance of trouble" (Ogden and Richards, 1923, p. 231).

I'd like to avoid trouble and get a grip on what 'expression' means, while acknowledging that it's a slippery concept. Part of what makes the term so bewildering is that it is used in a wide variety of situations (Greene, 2007, p. 21). Consider:

(a) The abnormal gene expressed itself in the cell.
(b) Cynthia's expressed desire is not to be bothered while she is in the meeting.
(c) Arthur's sign expressed frustration.
(d) This painting expresses anguish.
(e) The dog's barking expressed danger.
(f) Rose expressed her compassion by writing a note.
(g) Jean's career choice expresses a commitment to serving others.
(h) Jessica's facial expression expresses indifference.

When I say that actions express our selves, I don't have in mind uses of express like those in (a), (d), or (e). In each of these, it is not a person, but a thing (or unthinking animal) that expresses. I also don't have in mind (b), (c), or (h). Expressed is used as an adjective in (b) to emphasize Cynthia's desire, to indicate that she not only has a particular desire but previously shared it with

her colleagues so as to emphasize how important it is that she be left alone. We usually take sighs or facial expressions, as in (c) or (h), as signals of a person's emotional state. This is because certain sounds and facial expressions typically indicate certain emotions or moods. But in making such a sound or having such a facial expression, one need not be *self*-expressing anything. Self-expression and 'expressiveness' are, thus, distinct (Greene, 2007, p. 40). Facial expressions, sighs, body language and similar kinds of phenomena may be typically expressive of a particular emotion or feeling without expressing something about the person who sighs, uses particular body language, etc. in every instance, as, for example, when someone is acting. Of course, they may also genuinely show what they purport to express, as when Arthur is enraged, or Jessica actually indifferent. But as written in (c) and (h), 'express' is ambiguous between 'is expressive of' and the more technical definition of 'express' or, even more precisely, 'self-express' that I am after.

I will take as paradigm cases of expression in instances like (f) and (g). These cases are unique because what is expressed isn't ideas or concepts or emotions, but something about Rose and Jean, respectively. Rose's note wasn't merely describing compassion, it expressed *her* compassion. Jean's career choices revealed something about who she is and what is important to her. Our question, then, is: what does it mean to say that actions express something about the actor? What does expression mean here? Since we're concerned with expression of something about the expresser herself, these are also cases of self-expression. I will lay out what I take to be central features of self-expression.[30]

[30] The five features of self-expression that I lay out in this section are informed by Greene's (2007) twenty dicta about self-expression. See especially Chapter Two of Greene (2007). Greene's analysis of self-expression concerns all types of expression – in language, facial expression, embodied expression, and animal expression. I'm concerned with self-expression in a narrower context, namely the context of actions. I've combined some of his twenty dicta and distilled what I take to be relevant to my project. The view of self-expression that I lay out is not the same as Greene's view, though I've noted in the text where they overlap.

6.1.1.1 Self-Expression Shows Cares

In this section, I will unpack, in order, what it means to 'show' something and what 'cares' refers to. There are three ways in which certain actions, emotions, reflexes, or movements – self-expressions – might be thought to show another thing:[31]

- Showing that: Here one makes available knowledge of proposition. (I show you that I like *Downton Abbey* by watching *Downton Abbey*.)
- Showing-Y (where Y is a perceptible object): Here one makes an object perceptible. (I show you my scar by rolling up my sleeve and presenting you with my scar.)
- Showing how some emotion or experience feels or appears: Here I do something that puts you in a position to know what an experience of mine is like. (I might describe a recent meal in way that allows you to imagine you were there; I might write a note to a loved one trying to capture what it means to love her.)[32]

When activities, like those above, are designed to convey, typically mean to convey, or chosen to convey, the information that they do, they also show something. Self-expressions may show aspects of one's self in any of these three ways.

Now that we have an idea what it might mean to "show" something, we can turn to what kinds of things self-expression might show? The obvious answer is that self-expression shows the self. We show our selves by making available propositions about our selves, making our selves perceptible to others, or showing how some emotion or experience feels. But what is the self? What are we showing?

This is such a loaded question; I cannot possibly to justice to the many of theories of the self in philosophical literature.[33] I will assume that a particular position is correct, while acknowledging controversy. A variety of theories of the self (sometimes called "deep" self) have

[31] For those who are more familiar with philosophy of language and models of communication than I am, it may be useful to point out that showing is a type of signaling. A signal is any cue that was designed to convey the information that it does (Greene 2007, p. 5). The notion of design may include, but is not limited to, human intention. A biological signal may be a product of evolution, and so designed in the relevant sense.
[32] Greene (2010) distinguishes clearly between these three kinds of showings.
[33] For a recent review see Gallagher (2011).

been put forward, and these theories differ in important ways. Nonetheless, all theories of the self share the view that, of the totality of attitudes in a person's psychology, there is a distinguished subset of them that are fundamental to her practical identity. These attitudes, dubbed the ''self,'' belong to her in a distinctive way that carries significance for a number of aspects of agency (Sripada, 2015). The criteria used to pick out this subset of attitudes vary from theory to theory.

Theorists have tended to favor cognitive views that understand the self in terms of rationally formed subset of beliefs about oneself or evaluative judgments. An example is Frankfurt's account of the self. Frankfurt (1971) suggests that desires are part of the self when the subject reflectively endorses them as determinants of her behavior. Frankfurt explains: "It is only if [the agent] does want to X that [she] can coherently want the desire to X not merely to be one of [her] desires, but, more decisively, to be [her] will" (p. 10). Frankfurt's view is a cognitive view of the self because he understands the self as only those aspects of her psychology – only those desires – that are "coherently" wanted and reflectively endorsed. In contrast, Sripada (2015) proposes instead a conative view that says one's (deep) self consists of one's cares. They need not be reflectively endorsed or rationally coherent. According to Sripada, "cares involve a complex syndrome of motivational, commitmental, evaluative, and affective dispositions" (p. 9). I will review these four aspects of cares.

Sripada says the motivational effects of caring mean that caring about something always serves as the intrinsic motivation for actions that promote the achievement of that thing. In other words, the set of cares that make up the self are not instrumental desires; cares are distinctive in lying at the foundations of this hierarchy of motives (p. 6). For example, Katya wants to get the bus, in order to get to class, so that she graduates from university, all in service of becoming a

competent trader. Katya wants to become a trader because she cares about earning a lot of money. Cares provide motivational support for elements of one's subjective motivational set.

Cares also have commitmental effects. In caring about X, a person is not only intrinsically motivated by X, but also committed to continue to be motivated by X. Sripada says,

> The commitmental features of caring can be brought out more vividly by considering a person's responses to the prospect of changes to the elements of her conative set. If a person's pro-attitude towards X is merely a desire, and in particular a desire that is not in any way instrumental to anything she cares about, then she should be relatively indifferent to the prospect of this attitude being altered in some way: for example, being replaced by some attitude Y (where Y too is similarly irrelevant to her cares). In contrast, if she cares for X, then it would strike us as strange if she were indifferent to the prospect of change—if offered a pill that would erase one of her cares, she says, ''Meh, doesn't matter to me if I take this pill. Either way.'" (p. 7).

Sripada is quick to point out that this does not mean that cares cannot be altered, or that they are a static set. Cares shift and change over time, as when, for example, getting older slowly leads to caring less about going out with friends on Friday night and more about taking full advantage of one's Saturday. The commitmental effects of cares make it that even when a person genuinely judges that some care must, all things considered, be changed or erased, if the care is really part of her self, the prospect of changing will not be viewed only positively. I appreciate my Saturdays, but reflect nostalgically on fun Friday nights out. Erasing cares will tend to be accompanied by an experience of loss. We are often reticent, if not unwilling, to change our cares because of their commitmental effects.

Third, Sripada says, cares have evaluative effects. A person who cares about something is disposed to form judgments about that thing and cast it in a positive light. For example, a person who cares deeply making a lot of money is disposed to judge that maximizing profits is valuable, and that the actions that maximize profits are valuable.

Finally, cares are associated with emotional responses that can shape and refine one's overall set of cares. Sripada provides the case of Paul to elucidate:

> Suppose Paul cares about the plight of children dispossessed by war in Sudan. If a hostile United Nations resolution on Sudan is forthcoming, Paul is disposed to a suite of ''signaling'' emotions such as anxiety and fear that concentrate his attention on the looming threat and give it precedence over other considerations. If Sudanese children are benefited or advanced in some way, Paul is disposed to a suite of positively valenced emotions such as joy, approval, and elevation. If the fortunes of the Sudanese children are set back, Paul is susceptible to sadness, disapprobation, and despair. In this respect too, cares are quite different than desires. It is perfectly possible to desire something, but not have this rich and distinctive profile of emotional connections to the prospect of that thing being threatened, achieved, or foreclosed (p. 8-9).

Emotional responses reveal our cares to ourselves and to others. In that way, our emotional responses afford us access to our cares in a way that can facilitate shaping and refining one's set of cares.[34]

Sripada ultimately argues that one's set of cares (directly) constitutes one's self.[35] One of the main reasons he provides for thinking this is correct is that the functional role of cares establishes what matters to us. When something matters to a person, she is motivated to bring it about, she is reticent to give it up, she is committed to it, and she feels positively toward that thing. Sripada says:

> It is simply not possible that these descriptions are all true of her and yet the thing in question is unimportant to her, or worse, that she is alienated from the thing. These observations suggest that there is a basic conceptual tie between the syndrome of dispositional effects associated with cares and what it is for something to matter to a person" (p. 9).

That our cares matter to us is evidence that they are constitutive of who we are.

---

[34] See also Betzler (2007).

[35] Sripada provides an account that links one's self to one's cares in a direct strategy. This differs from an indirect strategy of linking one's self to one's cares, as Jaworski (2007) advances. An indirect strategy says that cares support some property X, where X is not part of the syndrome that characterizes cares. It is then argued that any state that plays this role with respect to X is part of one's deep self. To defend this, Jaworski invokes Bratman's (2000) claim that any state that plays this role in sustaining cross-temporal continuities and connections must necessarily belong to the person in precisely the way that is characteristic of elements of the deep self. Sripada says that, "The direct and indirect strategies are interlinked and ultimately complementary" (2015, p. 9, footnote 15).

I think Sripada's notion of the self best aligns with what I take ourselves to be 'showing' via self-expressions. It is a minimalist conception of the self. Sripada does not introduce machinery that would make it the case that only *rational* cares can constitute one's self, or *reflectively endorsed* cares. This means that when we express our selves, we show our flaws and all. If a smoker, Jamie, constantly talks about how much she values her health and wants to quit smoking, saying that's "really who she is," yet we see her going out for a cigarette at lunch every day, we treat her reflectively endorsed desires somewhat skeptically. We might think that her going out for a cigarette also expresses something about her, even though it's a part of her self she may not like. If we are really to take seriously the idea that her real self is constituted only by reflectively endorsed desires, then we are not to ascribe anything to her self on the basis of her contrary actions. Her going out for a cigarette shows something, but not something about her self. It strikes me as odd to say such a thing. What else could it show except something about Jamie?

One might argue in response that we often do things that we don't take to be 'showing' what we care about. Jamie might have another cigarette even though she genuinely cares about her health and very much wants to quit smoking. It might seem natural to say that this behavior does not express her 'self' insofar as it doesn't express what really matters to her. Anderson's (1993) expressive theory of *rational* action seems to imply something like this. Anderson says that the rational act is one that adequately expresses the agent's rational attitudes towards people and other things that she intrinsically values (p. 17). She says, "something is valuable if and only if it is rational for someone to value it, to assume a favorable attitude toward it" (ibid.). Now, admittedly, Anderson's standard of rationality is pretty minimal. She allows that we may rationally value lots of different things in lots of different ways – as long as they meet "certain

general standards" (ibid., p. 19). Still, her account saddles rational requirements regarding what attitudes one ought to have onto a description of the relationship between expression and actions that self-express.

Sripada resists the move that says that one's self is limited to only those reflectively endorsed or rational attitudes. When it comes to the unwilling smoker, he says, one of two things may be going on. Which one is correct depends on what Jamie is like, who she is. The first is that the urge to have another cigarette arises without a basis in her cares. Compare someone who is afraid of ordinary spiders but sees her reactive behavior to spiders as completely foreign to herself. Things she cares about play no functional role in supporting or bringing about her spider-directed fear. In this case, Sripada says, the occurrence of spider-directed fear would not express her self. Jamie might be like the arachnophobe. If the desire to have a cigarette is not supported by her cares, then that urge is not expressive of her self. On the second interpretation, Sripada allows that it is entirely plausible that a desire for a cigarette is supported by Jamie's cares, even though she also cares about her health. Sripada advances a mosaic conception of the self, one that allows conflict within one's set of cares. A mosaic conception of the self contrasts with a homogenous conception, which says that elements of one's self cannot conflict with another; apparent conflict can always be resolved upon deeper reflection. A homogenous conception would say that even when it seems that two elements of one's self (whether evaluative judgments or conative states) conflict, it must be that one of them is not really part of the self. I agree with Sripada's announcement that selves are "complex, variegated things" (p. 26) and that our cares conflict (p. 24). This is a feature of the human condition, and the best account of the self should not seek to erase it. The addict may have cares that deeply, sometimes tragically conflict.

Wanting another cigarette may be expressive of one of her cares, while wanting not to want another cigarette is, too.

Now, armed with a working theory of the self, and an understanding of what it means to show the self, we can arrive at the conclusion that what it means for something to be a self-expression is for it to show cares – the set of conative attitudes that are constitutive of the self. Something 'shows' cares when it makes available propositions about cares (e.g. their content), makes our cares perceptible to others, or shows how it feels to care about some object, and it is designed or intended to do so. Before moving on to how actions specifically show cares, I must make a couple of other remarks about self-expression.

6.1.1.2 Self-Expression Shows *One's* Cares

When we self-express, whose cares do we show? I will not belabor the point, but I take it as a platitude that one can show only one's own cares, i.e. one's own self. One the face of it, we seem perfectly capable of feigning emotions, desire, or cares. One can express regret even though she is not regretful. However, one who does not harbor regret is doing something that is merely looks like regret without self-expressing regret. This may be an expression of some other care, and even chosen on that basis, e.g. for the sake of continuing a business partnership. But at that point, feigning regret merely looks like regret, but actually expresses something else, namely a desire to continue to do business, which may be rooted in a care to earn a lot of money. Dramatic performances are also not self-expressions, since the actor is performing actions and saying things that are expressive of another person while not expressing his self. Indeed, actors learn elaborate techniques to achieve emotional expressiveness without self-expressing.

6.1.1.3 Self-Expression Can be Overt or Non-Overt

Overt self-expression refers to cases where one explicitly intends to show something about her self when she performs some action or movement. Non-overt self-expression occurs when an action or movement expresses something, but the actor did not intend explicitly to express when she undertook the action (Greene, 2007, p. 32-3). An example is choosing what to eat at a restaurant. We choose what to eat usually based on taste, or what sounds good. We don't typically think, "What does this express about me?" when reading a menu and deciding what to order. One could think of example where expression of cares is explicit, such as eating locally sourced food to underscore one's commitment to growing the local economy and communicate that to colleagues. Most self-expression, though, is non-overt. As an empirical point, I doubt most people think about what their self-expressions show about themselves. Showing something about one's self need not be among the motives for action, facial expression, or other forms of self-expression, even though they are genuine self-expressions.

6.1.1.4 Self-Expression Does Not Require an Audience

In the same passage as quoted above, Ogden and Richards said, "What is wanted is a searching enquiry into the processes concealed by [the term expression]…the introduction of the listener does little to throw light upon the matter" (1923, p. 231). This is to say, self-expression does not require an audience. No one must "hear" the expression in order for it to be one (Greene 2007, p. 31). Consider what one chooses to do in the privacy of one's own home. Watching *The Real Housewives of New Jersey* instead of *PBS News Hour* surely shows something about the watcher, even though she watches alone.

6.1.1.5 Self-expression Can be Successful or Unsuccessful[36]

"Success" is measured by the extent to which what one 'shows' is supported by one's cares and associated elements of her motivational set. One would fail to self-express if, as in the arachnophobia case, her behavior was not supported by a particular care. Another obvious way to fail to self-express is when one is incapable of making anything public, as one's words stick in her throat or she is paralyzed. Or, one might perform an action incorrectly, as when a platonic hug turns mildly sexual because of an awkward mistaken hand placement. The huggee may interpret a sexual advance, but this is not something supported by the hugger's cares or subjective motivational set. The mistakenly placed hand rendered this self-expression is unsuccessful. Finally, one may believe that a particular action would express some element of her self, when, in fact in wouldn't. I may believe that drinking the glass in front of me shows that I want a gin and tonic, believing it to contain gin and tonic, even though the glass does not contain gin and tonic.

Self-expression can be successful or unsuccessful even one when is self-expressing non-overtly. In the hug example, one might have spontaneously hugged, not meaning to express a particular care, even though it is rooted in concern for a friend in distress. That the mistaken hand placement showed something else rendered the expression unsuccessful. We measure success not by whether what was intended was successfully executed, but rather by to what extent what was shown was supported by the contents of one's self. Cultural convention will often play a role in determining success. We have typical ways of 'showing' concern for others, respect for authority, and delight in particular activity. What successful self-expression looks like in one time or place may not look like successful self-expression in another time or place.

[36] Cf. Greene (2007), "Like Other Acts, Attempts at Self-Expression may or may not be Successful" (p. 36)

6.1.2 Actions Self-Express

Now that we're getting a handle on what self-expression is, we can move to think about the connection between self-expression and actions. I laid out above a view whereby self-expression shows one's cares and associated elements of her subjective motivational set, e.g. the desires, evaluative judgments, and pro-attitudes that are motivational, commitmental, dispositional, and emotional effects of her cares. Self-expression can be overt or non-overt, does not require an audience, and can be successful or unsuccessful. Our current question is: do actions do this? Do actions show one's cares and associated elements of her subjective motivational set in the ways that I outlined above? To be clear, in this section what I have in mind are intentional actions, as opposed to mere movements (e.g. finger tapping) or facial expressions.[37]

Recall that one might show her cares in one of three ways: (1) showing that; (2) showing-Y; and (3) showing how. The first way of 'showing,' showing that, makes available knowledge of proposition. In the context of self-expression, it might show that one has a particular care, e.g. the proposition that one cares about her partner. Actions of course do this. It is not surprising to hear (or to say), "Don't tell me that you love me. Show me!" Unwillingness to do so is often a painful indication that one's love is unrequited.

The second way of showing, showing-Y, makes an object Y perceptible. When it comes to self-expression, that object is our self. Does action make our self perceptible? The answer, again, is of course. Action is not the only way in which we make our cares and associated

[37] I want to remain as agnostic as possible on what intention means in this context. I have in mind a folk conception of intentional action, paradigm case where an intention to act precedes the action itself, making no claim about what kind of mental attitude is an intention, whether intentions cause an action, whether intentions are explanations, whether intentions are (explanatory) reasons, or other related debates. For a review, see Wilson & Shpall (2012). Habits like brushing one's teeth or pouring cream in one's coffee can also count as intentional action, as far as I'm concerned.

subjective motivational set perceptible to others, but it is one of the most often utilized ways in which we do so. No action could purport to show one's entire self, but we do show parts of our selves in actions, thereby making perceptible parts of our self.

Finally, self-expression might show how some emotion or experience feels or appears to me. Can actions do that? This one may not be so obvious, but they can indeed show how some experience feels. Think about our use of adverbs to describe actions. She *tenderly* cared for her child. He *carefully* held his granddaughter. The doctor looked at her *inquisitively*. He *quickly* locked the door that led outside. In each sentence, one gets a sense of not only of what the subject is doing, but *how*. We wouldn't need adverbs if actions didn't also make perceptible to others how some action feels from the inside.

When actions show that I have a particular care, or how that care feels to me, they are self-expressions. When we combine that claim with Sripada's account of the self, we end with the statement that actions express one's cares and their associated motivational, dispositional, commitmental, and emotional effects. It does not follow from this statement that that all actions are self-expressions. Some actions, e.g. jumping back from a spider, might be triggered by something outside of myself, which is to say, unsupported by one of my cares and associated subjective motivational set. Some actions may be unsuccessful insofar as they are believed to show one's care, but do not in the context, or one misfires in acting, e.g. by putting one's hand in an awkward place by mistake. The arachnophobia case is different than cases of false beliefs or mistakes in acting. In the arachnophobia case, one is compelled to act by alien forces, let's call them. The action originates somewhere other than in one's self. In the latter cases, action originates in the self, but fails to express the self because of a mistake, either a mistaken belief or a misfire in acting.

The distinction between these two ways of being unsuccessful at self-expression yields an important insight about the connection between action and self-expression. Except when compelled by "alien desire" all action is rooted in cares and associated motivations, evaluative judgments, attitudes, desires, etc., even when it is ultimately unsuccessful at expressing those cares. There can be no other basis for action, insofar as there is no other way for actions to be unsuccessful at self-expression. Either an action fails to self-express because it has no support in the self, or it has support in the self, but fails to show the self.

One might want to allow that there are other bases for acting. Duty comes to mind. We might consider acting only because one believes that she has a duty to do so. When we say that one acts on the basis of duty alone, one is doing things because she has to, not because she cares about doing her duty. I think this statement is true, but it does not undermine the statement that actions self-express. Duties may be interpreted as "alien" forces, much like the arachnophobia case. Alternatively, one may desire to do one's duty, if only because one cares about pleasing authoritative figures, and so performing actions that fulfill one's duties genuinely self-express. The correct interpretation will depend on the person and context.

6.1.3 Actions, Reasons, and Self-Expression

I argued above that actions self express, and, further, that when they are unsuccessful at self-expression, they fail in one of two ways. Either "alien" forces trigger them, or they are supported by cares but fail in execution due to some mistake. What does this have to do with reasons?

Insofar as reasons prescribe actions, they prescribe actions that, if performed, may or may not successfully self-express. If they prescribe actions that, when performed, successfully self-

express, then we need not worry about what acting in accordance with reasons means for self-expression. I worry, however, when acting in accordance with reasons prescribes actions that are not self-expressive.

Consider what would be entailed by the supposition that one has a reason to ϕ, whatever else one wants, and that one could come to accept this conclusion absent information relation between ϕ-ing and her subjective cares (i.e. absent information about to what extent ϕ-ing is rooted in her cares). Let's say, for illustrative purposes that ϕ-ing is not supported by my cares or subjective motivational set, yet I do it anyway, since one claims that I have a reason to ϕ.

In this case, ϕ-ing would amount to my play-acting as someone I am not, acting as if compelled by alien forces, much like the arachnophobe. Just as jumping back from a spider feels like it's "not me," so, too, does ϕ-ing.

I have three worries about the idea that we have external reasons, at least some of the time, to perform actions that do not express our cares, and would not even after sound deliberation. First, when reasons indicate that I perform actions that do not express my cares in this sense, acting accordingly is not an expression of *my* agency. It is unclear how acting in accordance with the reason to perform that action would be an expression of me. Williams, pointed out this problem in his indictment of Utilitarianism (1973). Against utilitarianism, Williams argued:

> 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 demanded, that he should just step aside from his own projects and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his acts and the source of his actions in his own convictions (in Smart & Williams 1973, p. 116-7).

The absurdity consists in having reasons to perform actions that could not possibly count as actions performed *by me* insofar as an agent with this particular set of cares would not be

motivated to act in that way. The same absurdity arises whether we are dealing with utilitarian reasons, objectivist reasons, or certain subjectivist-externalist reasons.

Second, acting on "alien" forces disrupts the process of self-development. As alien actions, like jumping back from a spider, do not originate in the self, acting on them does not support self-development. If anything, it impedes self-development by entirely skirting around the self. Self-development occurs by exploring, honing, and finessing one's cares by cultivating the talents, abilities, capacities, desires, evaluative attitudes, and beliefs that those cares support. It is an active process, contrasted with passive consumption. Self-development has long been held to be independently valuable. Mill quotes Wilhelm von Humboldt at the beginning of *On Liberty:*

> The grand, leading principle, towards which every argument, unfolded in these pages directly converges, is the absolute and essential importance of human development in its richest diversity (von Humboldt in Mill 1859/2011, p. 633).

Mill later says, in the same essay:

> He who lets the world, or his own portion of it, choose his plan of life for him has no need of any other faculty than the ape-like one of imitation He who chooses his plan for himself employs all his faculties. He must use observation to see, reasoning and judgment to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold his deliberate decision…It is possible that he might be guided in some good path, and kept out of harm's way, without any of these things. But what will be his comparative worth as a human being? It really is of importance, not only what men do, but also what manner of men they are that do it (Mill 1859/2011, p. 658-9).

Acting on alien forces is letting the world choose what one does. We may be kept out of harm's way and guided on a good path, but as Mill rightly points out, merely acting is not important. The tie between acting and "the manner of men" (and women) that act is what matters. That tie is self-expression.

Third, insisting that we have reasons to act as if by alien force risks undermining intersubjective experience. It not only alienates agents from their cares, from their self, but also

agents from each other. It creates a chasm between who we are and what we present to the world. A notable example of when this may occur is when persons with homosexual desires are forced, often in the name of morality, to suppress their own cares and desires, and instead act as if they were heterosexual. Now, I don't want to imply that a homosexual person couldn't have an internal or subjective reason to suppress homosexual desires, if, for example, she values her relationship with her disapproving parents more than her sexual life. But then it is not "alien" forces compelling her action. These kinds of cases aside, when the person with homosexual desires acts as if she has desires that are alien to her, she not only denies a part of herself, but also puts distance between her partners and herself. She cannot be herself around them.

This leads to a related problem. Part of the way we learn about our selves and the world is in conversation with others. If others are not capable of accessing information about our selves, if, e.g. we are repressing certain desires or aspects of our selves, they will be unable to provide useful information or feedback. When others provide advice or recommendations based on false beliefs about one's self, the other's advice and recommendations fail to connect with who one is, and what kinds of actions, experiences, choices, desires, etc. are in line with one's self. When we act as we aren't, thereby expressing alien desires, we undermine intersubjectivity, which is unfortunate given that intersubjectivity and interaction with others is a crucial part of learning about one's self, others, and the world around us. Manne (2014) was right that the recommendations and advice of others matters to reasoning practically or deliberating about our reasons, but those others must be able to access knowledge of our selves in order to have a successful conversation.

Certain externalists might resist this criticism. Proponents of IAA, for example, might insist that because, on their account, reasons needn't serve as the motive for acting, insofar as the

information required to determine what A has reason to do is not accessible to A, by default A will still in act in ways that are self-expressive. She just won't be acting in accordance with reasons. This glib response will not, and should not, satisfy someone who things that normative reasons ought, ideally, to guide action.

These three worries about requiring agents to act in ways that do not self-express motivate building a conception of reasons that avoids placing agents in that position. That our actions express who we are is important to protecting our agency, our selves, and our interpersonal relationships. So, rather than suppose that normative reasons prescribe actions, independently of whether those actions express our selves, it is reasonable to place a restriction on what can be our reasons for action. That restriction will say that we can have a reason to perform only those actions that are genuinely expressive of our cares and associated elements of our subjective motivational sets. What we can express about our selves is limited at the same time by the contents of our cares. This does not imply that our cares are fixed, only that at the time at which we act, there is a truth of the matter about what actions are genuinely self-expressions.

If one accepts this the connection between self-expression and actions, then one should also accept that the reasons that guide our actions ought to permit genuine self-expression. There are all sorts of norms that might govern behavior (some of them moral norms). My argument concludes that those norms are normative reasons for action for a particular agent when the actions they prescribe express the cares of the actor. This is the explanation constraint under a slightly different guise. It says not that reasons must be able to serve as the motive for actions in a strict sense, but rather that reasons prescribe actions that, if performed by the actor, would truly express her self. Since her self is a set of cares that have motivational effect, whatever actions

express one's self will be supported by those motivational effects. In this way, it falls out, as a contingent fact, that whatever one has reason to do, one will be motivated to do. Thus, this argument vindicates not only a version of the explanation constraint on reasons, but also S-I about reasons.

6.1.4 Advantages of the Argument for S-I from Self-Expression

The argument for S-I from self-expression does not assume the truth of Humean theories of motivation. Humean theories of motivation assert that *only* desires are capable of motivating action; beliefs alone will not suffice. A typical "Humean" theory of reasons will endorse some version of internalism on the basis that beliefs alone (e.g. beliefs about what we have reason to do) cannot motivate. The Humean theory of reasons is rooted in a Humean theory of motivation. There must be some basis for acting *in* the person, some desire that would spur action. While I am sympathetic to the idea that there must be something *in* the agent that grounds the truth of reasons-statements, the argument presented here is agnostic on whether beliefs, desires, or some other entity altogether serves, ultimately, as the motive for action. The entity that motivates action – a belief, a desire, or some other mental state – is irrelevant.

This is an important point because one of the most prominent objections to S-I is that S-I is committed to the claim that we cannot have reason to do that which we are not currently motivated to do. But S-I is not committed to that claim. S-I can accommodate the intuition that we often have reasons to perform actions that we are not currently motivated to perform. Most days, I am not motivated to wear my helmet when I bike ride. But I may have a reason to do so nonetheless, since helmet wearing expresses my commitment to the success of public health interventions. If I soundly deliberated, I would be motivated (at least in part) to wear a helmet. S-

I can, once again, be agnostic as to whether the agent is actually motivated by the belief that something is expressive if her self or whether more along the lines of a care, a desire, or emotion does the motivating.

Along the same lines, the argument does not rest on the truth of motivational internalism about reasons to defend its conclusion. Motivational internalism is the view that an agent would be motivated to ϕ upon recognizing that she has a reason to ϕ.[38] While this may be turn out to be true on any S-I account of reasons, it is not part of the argument for the explanation constraint. We can defend S-I about reasons without assuming motivational internalism. The problem with assuming motivational internalism is that one must also assume that a Humean theory of motivation is correct in order to reach a conclusion about S-I. This invokes two problematic theses about motivation: first, that one will be motivated upon realizing that one has a reason; and second, that motivational states do not include beliefs. For this reason, I think the argument provided is preferable to Kate Manne's (2014) argument for subjectivist-internalism. Manne's argument depends, as a premise, on the truth of motivational internalism.

While I don't need motivational internalism to defend a revised version of the explanation constraint, in fact I think it is likely a result of the argument, an entailment of the view. Expressions of self are linked to cares, which are motivational states. If we limit what one has reason to do to what is expressive of one's cares, one would be motivated, at least in part, to do whatever one, in fact, has reason to do, at least under ordinary conditions. This point about motivational internalism, however, is not necessary for the argument. Further work will defend the thesis motivational internalism is an entailment of the argument for S-I from self-expression.

[38] See Chapter One, Section 1.1.2, where I lay out a variety of internalisms.

6.2 Expressive Reasons

Having defended the idea that the connection between self-expression and action secures subjectivist-internalism about reasons, finally I will formulate my account of reasons, called "Expressive Reasons" (ER). Invoking Rosati's (1996) formulation of "two-tier" internalism, I propose two conditions on some consideration's having the property of being a reason.

> ER: R is a reason for A to ϕ when ϕ-ing is an expression of soundly deliberating A's self; and A, under ordinary conditions, would act on the basis of R.

The first clause of ER emphasizes the link between actions, reasons, and self-expression. The second clause makes explicit the connection between reasons and motives. Soundly deliberating A, or A+, is A, but under improved epistemic circumstances. I will spell out sound deliberation in the same way that Williams did, as explicated in the previous chapter. A has reason to do what A+ determines is an expression of her, which is to say, A's, self. The second clause of ER ensures that whatever A has reason to do will also be able to serve as the motive for acting, at least under ordinary conditions.

6.2.1 Sound Deliberation

My challenge in explicating sound deliberation is to retain the normative constraint on reasons, but to suitably idealize agents in a way such that agents would still be motivated act on the basis of those reasons under ordinary conditions. A doesn't have a reason to perform any act that expresses her current motivational set, but only those actions which would still express her self if she were soundly deliberating. Yet, ER also says that whatever A has reason to do, it must be the case that she would be motivated to act accordingly under ordinary optimal conditions. What does soundly deliberating A+ have that A lacks when she determines that ϕ-ing is an expression of A's self under the circumstances?

I favor Williams' account of sound deliberation. In respecting the explanation constraint upon reasons, Williams' account of sound deliberation also respects an important desideratum of my theory. Sound deliberation will not render A+ an unrecognizable person, such that she will recommend that A act in ways that are not supported by her self. Soundly deliberating A+ will have the same *cares* as A, even if she has more information about A's circumstances, available actions, and even A herself. ER will produce the right answers on some of the most oft-cited cases. For example, if A wants to have a gin and tonic, drinking a glass of petrol is not expressive of her desire. Soundly deliberating A will realize that, and request a new gin-and-tonic, instead. If A cares about spending time with friends, soundly deliberating A will realize that an evening at a quiet restaurant is a better expression of her self than going to a rowdy concert. A will, thus, have a reason to drink the glass of gin (or, at least avoid the glass of petrol) and a reason to go to a quiet dinner, respectively. She will not have a reason to drink the petrol or attend the rowdy concert.

Recall that sound deliberation requires more than just reasoning instrumentally from point A to point B. Often, the most difficult part about figuring out what one should do is not figuring out how to satisfy one's desires, or the series of steps to manifest one's care for others, it's how to do so in a way that is practicable, pleasing, convenient, efficient, economical, etc. In other words, it's figuring out how to satisfy one's most pressing care or concern amidst a whole host of other cares and concerns. Williams argued that A+ does more than reason instrumentally. A+ also has improved information and uses imagination. On the first of these, A+ has more information than A about her circumstances, about what actions are available to her, about others and the world, and anything else relevant to make a determination about what expresses A's self. Like Mason (2006), I interpret Williams as saying that A+ has improved information to the

extent that is possible without requiring substantial revision of one's self – her cares and associated elements of her subjective motivational set. Sound deliberation will not require us to give up cares with insurmountably strong commitmental effects. Sound deliberation requires that we update information that bears on what reasons we have only when doing is not inconsistent with other cares. We are sometimes "blocked" from accessing and appreciating even true information because it conflicts with our cares and associated elements of our subjective motivational set. An agent's epistemic circumstances and subjective motivational set affect what she can come to know.

Imagination enters into Williams' account of sound deliberation, too. Williams' inclusion of imagination as a part of sound deliberation permits that deliberation about how to satisfy desires and what desires we prioritize can take other more amorphous forms. In order to assess whether some consideration is a reason, we should think about what it would be like to act on that reason. Through reading books, watching movies, or conversation with others (among lots of other activities) we may come to have "some concrete sense" of what would be involved in satisfying some element of one's motivational set or acting on the basis of a particular reason (1981, p. 105). Williams continues that in imagining an agent may, "lose his desire for [some outcome], just as, positively, the imagination can create new possibilities and new desires" (ibid.). I refer the reader back to Chapter Five, Section 4.3 for a full explication of the conditions of sound deliberation.

6.2.2 Ordinary Conditions

Soundly deliberating A+ determines what actions would express A's self. I already argued that A will always be motivated to act, at least in part, in ways that self-express. The

second condition upon reasons makes explicit the tie between expressive reasons and motivation: A would ϕ on the basis of her reason, at least under ordinary conditions. The caveat that A need only be motivated to act for her reasons "under ordinary conditions" may seem problematic, for reasons mentioned in Chapter Five when I discussed Michael Brady's account of reasons. In saying that A need only be motivated to act under "ordinary conditions" reasons may be unable to motivate her in actual, suboptimal conditions. If ER allows that possibility, it will fail to be an internalist view.

The strength of this objection depends on how different "ordinary conditions" are from one's every day life. When she sets up her internalist account of the good, Rosati (1996) says that ordinary conditions "are simply those that we already accept as the minimal conditions that must be met for a person to think sensibly about her good at all" including, for example, "that a person not be sleeping, drugged, or hypnotized…whatever normally attainable conditions are optimal for reflecting on questions about what to care about" (p. 305).[39] I would add not physically constrained to this list. Someone may have a reason to escape a captor, but is incapable because she is tied to a chair. People who are drugged, hypnotized, physically constrained, enraged, or in the throes of some other mind-altering state are, typically, not able to respond to reasons. An internalist theory of reasons need not stipulate that reasons would motivate under those conditions. At the same time "ordinary conditions" is not meant to capture any more than a cool hour of deliberating, keeping the individual's cares, subjective motivational set and circumstances the same. Under these circumstances, ER says, one would be motivated to act in accordance with her reasons. Reasons are established by what her soundly deliberating self determines would express her self.

[39] Rosati inserts an "optimal" into "ordinary conditions," calling them "ordinary optimal conditions" instead. I've dropped the "optimal" because I think it unnecessarily makes it sound harder to achieve those conditions than is actually the case. Ordinary conditions really are most ordinary.

6.3 Some Objections and Replies

6.3.1 A+ is Not Idealized Enough

Johnson (2003) objected to Van Roojen's "somewhat less idealized" account of reasons on the basis that, "if we hold fixed all of the grounds of this reason, we do not idealize…at all" (p. 579). As a reminder, Van Roojen's proposal said: An agent has a reason to do an action on certain grounds and in certain conditions only if, of the complete descriptions of that agent that include those grounds and circumstances, the ones that make the agent out to be most rational and relevantly informed include a motive to do that action (Van Roojen 2000, p. 238). Van Roojen's proposal allows "somewhat less" idealization than competing views, which require full information or perfect procedural reasoning. It says that the idealization process holds fixed the agent's beliefs, desires, goals, and intentions (ibid.). Whatever rationality-maximizing actions she would still be motivated to do establishes what her non-ideal counterpart has reason to do. This presumably, is something her non-ideal counterpart would do, too.

The same criticism may hold against my view. Holding fixed A's cares (and the associated blocks to information) in the process of sound deliberation, A+ may not be that much better off, epistemically speaking, than A. A+ may not be able to determine which actions are genuine self-expression of A's self. But if we don't idealize A+, we lose the normative constraint upon reasons.

Weakening the requirements of the "idealization process," as I have in endorsing Williams' account of sound deliberation, may shorten the A+'s critical distance from A, but it does not eliminate it altogether. It will still be possible for A to fail to act in accordance with what she has reason to do, insofar as A will sometimes, if not often, fail to stop and reflect on her

circumstances before she acts. She will act impulsively. She will have mistaken beliefs and fail to imagine what something really is like. There is still a lot of room for improvement.

6.3.2 The Problem of Self-Knowledge

A+ will have not only improved information about the external world, but also improved information about her self. This information is crucial to determining what actions would express A's self. I make this statement knowing full well that we often fail to know our selves. This objection to ER says that insofar as we can't really know our selves, our soundly deliberating counterparts still will not be able to determine which actions are expressions of our selves.

A wealth of empirical work in psychology supports skepticism about the whether we can really know our selves and doubt about the reliability of introspection. Wilson and Dunn (2004) provide a lengthy review of literature that explores our limits to self-knowledge. Barriers to self-knowledge include repression, suppression, intentional forgetting, and complete forgetting. On this they conclude, "To the extent that people are motivated to block out thoughts, feelings, or memories, and succeed in doing so, self-knowledge will obviously suffer" (p. 499). But despite imposing limits to self-knowledge, self-expression may not "suffer" from these kinds of psychological machinery. For example, successful repression of painful memories may instead be the only way in which self-expression is possible, insofar as one may see the barrage of painful memories as an alien force. A more pervasive limit on self-knowledge, Wilson and Dunn argue, is the fact that, "much of the mind is inaccessible to conscious awareness (ibid.). Particularly recalcitrant parts of the mind include implicit perception, implicit biases, implicit learning, and implicit attitudes, among other implicit mental processes (ibid., pp. 499-504). Implicit processes have the cumulative effect of making it seem like we know what we're doing

and why, when, it turns out, we don't. We may question the extent to which implicit attitudes are part of our self, especially, e.g. racist implicit biases. Answering this in a way consistent with Sripada's account of the self requires asking whether implicit biases are supported by our cares, and to what extent we can accurately assess whether implicit biases are supported by our cares. Recall that our cares can include flaws as well as high notes.

Two arguably effective ways of improving self-knowledge include looking at oneself through the eyes of another, and inferring nonconscious states from our behavior (ibid. p. 507-11). I don't want to imply that A+ has some super-human capacity of introspection, but she can, to the extent allowed by who she is, form the most objective view of herself as is possible. In this way, A+ treats herself to both an insider and outsider perspective. She gains an understanding of who she is by seeing what her behavior reveals, and compares that to what she discovers about herself through introspection. Since there is no consensus as to how to evaluate the results of introspection, however, or what weight to accord other sources of evidence about ourselves, e.g. behavior, it's impossible to determine which is the more reliable method of arriving at genuine self-knowledge. Moreover, since introspection would be required to calibrate any tool that attempted to measure the accuracy of self-knowledge, there seems to be no way of "calibrating" it (Goldman 2004, p. 14). The best A+ can do is look to different sources and put together a coherent picture. From there, she can determine which actions available to A express her self.

Notwithstanding challenges to self-knowledge, A+'s improved epistemic state will provide a vantage from which she is in a better position that A – using instrumental reasoning, improved information, and imagination – to make a determination about what actions are self-expression of A in that particular time and place. The fact that A may not arrive at the same conclusion in practice (being in a less ideal epistemic position) does not vitiate that there is a

truth of the matter as to what actions express her self in the circumstances. It just means that sometimes A will be wrong about what she has reason to do. Nonetheless, whatever A has a reason to do would be able to serve as the basis for her acting under even under ordinary circumstances, given that a sound deliberative route from A to A+ exists.

6.3.3 ER is Incoherent

This objection to ER says that it is incoherent. It attempts to undermine ER by arguing that ER assumes that agents have at least one objective reason, namely a reason to self-express. If successful, this objection undermines ER's subjectivism.

Obviously, I'd like to resist this attack. While I am committed to thinking, with Mill, that cultivating and acting as one's self is an essential part of being human, I don't need to assume that we have a reason, on that basis alone, to do so. Normative reasons do not derive their normativity from being expressions of one's self, per se. The source of normativity is one's cares. What we do does not derive its value simply from being a self-expression. Its value comes from being connected to something that we care about.

That cultivating and acting as one's self is an essential part of living is important to defending ER, but it is not essential to establishing the truth in practice about what considerations have the property of being a reason. Self-expression need not be antecedently desired for its own sake by sound deliberators. That is, we need not be concerned with developing a self and we need not prioritize that above other values in the process of self-development. We develop a self and express that self regardless of whether we overtly mean to do so. So, though I must assume that it is true that self-expression is an essential and valuable part of human experience, I need not assume that we all share a reason to self-express, since we

need not care about our self or self-expression in order to achieve self-expression. One self-expresses whether she cares about it or not. Since one may lack a care for self and self-expression, there may be no reason to self-express, according to ER. Still, one may have reasons to perform lots of other actions rooted in what she does care about. I need not claim that we have reasons express our selves that ground our reasons to perform actions that are self-expressive. We therefore do not need to the claim that we have antecedent reason to self-express.

Relatedly, one may ask: why act in accordance with reasons at all? I don't have an answer to this question. This is no more of a problem for my view than it is for competing accounts of reasons.

6.3.4 Certain Selves Shouldn't be Expressed

There are two oft-cited example of the kinds of things no one could possibly have reason to do: count blades of grass (Rawls 1971) and hurt one's spouse (Williams 1981, 1995a). Though motivated by different concerns, they both amount to saying certain selves or certain aspects our selves should not be expressed. The objection says that any theory that leaves open the possibility that one has a reason to count the blades of grass in Central Park or hurt his spouse must be flawed. I'll take the examples in turn.

When one asserts that one could not possibly have a reason to count the blades of grass in Central Park, one usually assumes something like a rational person would not choose that activity. It would be a perversion of human rationality and human faculties to opt to count blades of grass when one could, instead, engage in conversation, learn, read a book, or do almost anything else except count blades of grass. It is irrational to care about counting to very high numbers for its own sake.

While I admit that it would indeed be odd to care so deeply about the number of blades of grass in Central Park, there is nothing in my account of reasons that says one could not have a reason to perform such an action. According to ER, if counting blades of grass is, upon soundly deliberating, determined to express one's self, perhaps on the basis that one derives the highest pleasure from counting to very high numbers, then one has a reason to do so. The person who has this reason may be odd, but she does not necessarily lack a reason for her actions. I take as a virtue of my view that it allows a wide diversity in what we care about and in how we shape our lives. Admitting that someone, some place might have a reason to count blades of grass is a small price to pay to maintain the conclusion that what we have reason to do must be an expression of who we are. In very rare cases, that might be counting blades of grass.

In admitting that someone, some place might have a reason to count blades of grass, must I also let in that someone, some place could have a reason to hurt his spouse? One can rely on one of (at least) two claims to argue that the husband does not have a reason to abuse his wife. One can focus on the husband, and insist that nature of rationality renders it is irrational for a husband to abuse his wife, or contrary to leading a good life. Or, one can claim that something about his wife provides a reason for him not to hit her. Both of these strategies invoke an objectivist conception of reasons. Either some substantive standard of rationality makes it such that no rational or good person would be motivated to perform this act, or something about the value of persons means that one has a reason not to abuse them. Either way, no matter how much the husband wants to hit his wife or cares about hurting her, he has no reason to do so.

My response to this line of thinking is the same as my response to the grass-counter. While I admit that it would indeed be odd to care so deeply about hurting another person, there is

nothing in my account that says one couldn't or shouldn't care about such a thing. I appreciate Williams' response to this case:

> There are many things I can say about or to this man: that he ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things. I shall presumably says, whatever else I say, that it would be better if he were nicer to her. There is one specific thing the external [or objective] reasons theorist wants me to say, that the man has a reason to be nicer…But if it is thought to be appropriate, what is supposed to make it appropriate, as opposed to (or in addition to) all those other things that may be said? The question is: what is the difference supposed to be between saying that the agent has a reason to act more considerately, and saying one of the many other things we can say to people whose behaviour does not accord with what we think it should be? As, for instance, it would be better if they acted otherwise (Williams, 1995a, p. 39-40).

Williams is asking, rhetorically: why must we insist this man is irrational in addition to all of the other things that the internal reasons theorist can already say? What does that charge add? Williams thinks it's very little: "we launch them in hope that somewhere in the agent is some motivation that by some deliberative route might issue in the action we seek" (ibid.). In other words, we launch whatever critical words we can in order to convince this person that he has an internal reason to do otherwise; perhaps that this isn't who he really is, or who he wants to be. Ideally, we would show the domestic abuser that given his set of cares, hurting his wife is not an expression of his self. It frustrates the intentions formed on the basis of other cares and associated desires, e.g. the desire not to go to jail, or his care to make one's parents proud. We may not want certain aspects of others' selves to be expressed because some cares are harmful or nefarious or selfish or brutal. But not wanting it to be the case does not establish the impossibility of having a reason to act in ways that express those cares.

6.4 Not Anything Goes

Recall Savulescu's assessment of subjectivism about reasons: "[I]f we happened not to care about human beings, or persons, we would have no reasons…to care about them. If parents

did not care about their children, then they would have no reasons to care about them." He concludes, "[A]nything goes…or at least anything could go depending on what we happened to care about" (Savulescu 2009, p. 225). Savulescu's statement captures the typical, visceral reaction to S-I theories of reasons. It's usually something like "You can't have a reason to do that!" or "But you *do* have a reason to do this!" Undergirding those statements is some species of objectivism about reasons and the assumption that there are universal reasons for all.

Objectivism endorses the existence of universal reasons, but it's important to point out that subjectivist views may well admit of reasons for all, even if the existence of reasons is contingent upon there being substantial overlap in what we care about. There is a difference between saying "there is a reason to X" and "there is some reason R that is a reason *for all* to X." The former implies that there exist reasons independently of any information about specific agents. I'll call this *necessary* universality. These are usually something like moral facts, values, or principles. The latter claims that all individual agents share some reasons, in so the reasons are in some sense universal, but does not imply the existence of these reasons independent of the agents themselves. I call this *contingent* universality. In Chapter Two, I denied the existence only of the *necessary* claim that posits the existence of universal reasons whether or not acting upon them is an expression of an agent's self or supported by what she cares about. The *contingent* claim – that there is a reason *for all* to X – might well be true, even on a subjectivist view. This is because our cares overlap, often substantially. That does not preclude an outlier, like our grass-counting friend or domestic abuser enemy may have a reason to do something that is, on first pass, completely crazy. But it does show that S-I can support the claim that, to the extent that our cares overlap, we will all share certain reasons. If we have some overlap in cares and desires, a discussion about how to be a parent, how to have meaningful relationships, how to be a decent

citizen, or how to be a supportive friend can ensue. A subjectivist about reasons can still say that some ways of being a parent, a citizen, or friend are better and worse without admitting that there is some objective point from which disagreement can be resolved. That is to say, the subjectivist can be critical of others, and wish that their reasons were different, without positing a set of objective reasons about which we cannot disagree. We can still seek, and achieve, agreement within a subjectivist framework of reasons, even if only contingently.

6.5 Conclusion

Expressive Reasons says that we have a reason to ϕ when ϕ-ing would express ourselves if we were soundly deliberating, and when we would act on the basis of the reason under ordinary conditions. This is a squarely subjectivist view. The source of normative reasons, what generates the reason in the first place, are the cares that constitute one's self. But in relying on a conative view of cares, permitting that our selves can be variegated, complex, sometimes contradictory sets of cares, I haven't further claimed that reasons are generated by the cares that are consistent with rationality, or cares that are conducive to leading a good life, or cares that one would have when viewing the world impartially. Sound deliberation does not require that much. Any care can, in principle, generate a reason, even cares that indicate nefarious, self-destructive, or terrible things, as long as one would persist in holding that acting on that care is expressive of one's self when soundly deliberating.

Why allow this? My strategy all along has been to provide philosophical reasons to reject objectivist and subjectivist-externalist views. Objectivist-externalists, I argued, are subject to an open-question type argument. We can always ask of a particular (objectivist) reason, but why that? Objectivist internalists face a dilemma: either they are also subject to the open-question argument, or their theory collapses into a subjectivist view. Finally, subjectivist-externalist views

permit alienation from one's reasons in affixing our reasons to a perspective unrecognizable and incomprehensible from the current first-person perspective. After this process of elimination, I concluded that subjectivist-internalism was the correct account of reasons. I argued that Expressive Reasons is the best S-I account of reasons, insofar as it does not assume motivational internalism and it best explains which cares, desires, concerns, and other elements of one's subjective motivational set generate reasons and which do not.

But for all the philosophical reasons to favor S-I over other families of reasons, I am attracted to it, too, for its practical pay offs. Because there is substantial overlap in what we care about, we share many reasons, including paradigm moral reasons, such as reasons to help strangers. We see so much overlap in what we care about because what we care about is a product of our social and cultural milieu. Where there is not overlap, our response should not be presumptuousness about what one ought to desire or ought to do, but rather inquisitiveness. What sorts of social structures and relationships give rise to a set of cares radically different than my own? What sorts of cultural norms or conditions contribute to scenario in which perpetrating gang violence, is, all things considered, the best expression of one's self? My goal is, and has been, to vindicate the individual in that situation – to recognize that one may be justified in acting despite being a place where what one has reason to do is something most of us wish did not express anyone's self. I think there's a practical and social pay off to being able to say that this person acts in accordance with what she genuinely had reason to do, rather than saying it was bad but excusable, horrific but not blameworthy, irrational but "made sense given the circumstances." She's doing something justified by who she is – what she cares about – even if the set of cares that make up her self is a product of deplorable social structures, cultural norms, and oppressive conditions. I think this is sometimes a tragedy. I wish the social structures,

cultural norms, and oppressive conditions in which one has a reason to participate in terrorist groups and activities, for example, were different. But I also want to vindicate individuals within those systems – to say that *she's* doing what she ought to do, what's right even, given who she is and the circumstances in which she finds her self. Subjectivist-internalism about reasons gives me a theoretical framework where that's possible.

While reasons exist relative to persons on this view, insofar as they are grounded in subjective cares and desires, moral considerations, like principles of equality or respect, can still be thought of as worthy ideals that we strive for. I like Lisa Tessman's recent account of worthy ideals: "ideals that we deem unattainable by worthy and that serve a non-action-guiding purpose" (2014, p. 199). She says that sometimes what is possible under oppressive conditions is *not good enough* (ibid., p. 7, emphasis original). I'd like to adopt and adapt that to say: sometimes, what an agent has normative reason to do is *not good enough*. We'd like it to be different, for people to act in ways that are more conducive to justice, flourishing, public health, or solidarity, among other aims. But we cannot ignore the social structures, cultural norms, oppressive conditions, and other kinds of institutions that form our selves or the connection between who we are – our practical identities and contingent cares - and what we have reason to do. We can, at the same, time vindicate people for acting in accordance with reasons that express who they are and still think that there are ideals out there worth striving for. Accounts of worthy ideals and normative theory more generally can help us deliberate about what we have reason to do and about how to lead our leaves. But it doesn't establish what we have reason to do. Our soundly deliberating selves do that.

References

Anderson, E. (1993). *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Anscombe, G.E.M. (1957). *Intention*. Oxford: Basil Blackwell.

Anscombe, G. E. M. (1983). 'The Causation of Action. In M.Geach and L.Gormally (Eds) *Human Life, Action and Ethics*. Exeter: Imprint Academic, 2005: 89–108.

Arkonovich, S. (2013). Varieties of Reasons/Motives Internalism. *Philosophy Compass* 8(3): 210-219.

Bagnoli, C. (2015). Constructivism in Metaethics. In E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy (Spring 2015 Edition) Retrieved from:* http://plato.stanford.edu/archives/spr2015/entries/constructivism-metaethics

Battin, M. (1999). High-Risk Religion: Christian Science and the Violation of Informed Consent. In P. DesAutels, M. Battin and L. May (Eds.), *Praying for a Cure: When Medical and Religious Practices Conflict* (pp. 7-36). Lanham, MD: Rowman & Littlefield Publishers.

Behrends, D. (2015). Problems and Solutions for a Hybrid Approach to Grounded Practical Normativity. *Canadian Journal of Philosophy* 45(2): 159-78.

Berg, A. (Director). (2015). *Prophet's Prey* [Motion picture]. United States: Imagine Entertainment.

Betzler, M. (2007). Making Sense of Actions Expressing Emotions. *Dialectica* 61(3): 447-466.

Blackburn, S. (2006). Antirealist Expressivism and Quasi-Realism. In David Copp (Ed.) *The Oxford Handbook of Ethical Theory* (pp. 146-162). Oxford: Oxford University Press.

Brady, M. (2000). How to Understand Internalism. *Philosophical Quarterly* 50(198): 91-97.

Brandt, R. (1979). *A Theory of the Good and the Right*. Oxford: Clarendon Press.

Bratman, M.E. (2000). Reflection, planning, and temporally extended agency. *Philosophical Review* 109(1): 35-61.

Burke, R. (1983). *Internalism, Externalism, and Reasons for Action*. Doctoral Dissertation: The City University of New York.

Chang, R. (2013). Grounding practical normativity: going hybrid. *Philosophical Studies* 164(1): 163-187.

Chang, R. (2014). Practical Reasons: The Problem of Gridlock. In Barry Dainton & Howard Robinson (Eds.) *The Bloomsbury Companion to Analytical Philosophy* (pp. 474-499).

London: Bloomsbury Press.

Collingwood, R. (1938). *The Principles of Art*. Oxford: Oxford University Press.

Crisp, R. (2005). Value, Reasons and the Structure of Justification: How to Avoid Passing the Buck. *Analysis* 65: 80-85.

Croce, B. (1901/1992). *The Aesthetics as the Science of Expression and of the Linguistic in General*. Trans. C. Lyas. Cambridge: Cambridge University Press.

Dancy, J. (2000), Practical Reality. Oxford: Oxford University Press.

Darwall, S. (1983). *Impartial Reason*. Ithaca, N.Y.: Cornell University Press.

Davidson, D. (1963). Actions, Reasons, and Causes. *Journal of Philosophy* 60 (23): 685-700.

Davis, W. (2003). *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.

Dreier, J. (2005). Moral Relativism and Moral Nihilism. In David Copp (Ed.) *The Oxford Handbook of Ethical Theory* (pp. 240-64). Oxford: Oxford University Press.

Dreier, J. (2015). Another World. In Robert Johnson & Michael Smith (Eds.) *Passions and Projections: Themes from the Philosophy of Simon Blackburn* (pp. 155-71). Oxford: Oxford University Press.

Ebels-Duggan, K. (2014, December 6). [Review of the book *Moral Reason*]. *Notre Dame Philosophical Reviews*. Retrieved from http://ndpr.nd.edu/news/54535-moral-reason/

Enoch, D. (2005). Why Idealize? *Ethics* 115: 759-87.

Falk, W.D. (1947). 'Ought' and Motivation. *Proceedings of the Aristotelian Society* 48: 111-138.

Finlay, S. (2009). The Obscurity of Internal Reasons. *Philosopher's Imprint* 9(7): 1-22.

Finlay, S. & Schroeder, M. (2015). Reasons for Action: Internal vs. External. In E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy (*Winter 2015 Edition). Retreived from: http://plato.stanford.edu/archives/win2015/entries/reasons-internal-external/

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1): 5-20.

Frith, R. (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research* 12(3): 317-345.

Gallagher, S. (Ed.). (2011). *The Oxford Handbook of the Self*. Oxford: Oxford University Press.

Garrard, E. & McNaughton, D. (1998). Mapping Moral Motivation. *Ethical Theory and Moral Practice* 1(1): 45-59.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Gert, J. (2002). Avoiding the Conditional Fallacy. *Philosophical Quarterly* 52(206): 88-95.

Gert, J. (2004). *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*: *A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gilbert, E. (2006). *Eat, Pray, Love*. London: Penguin Books.

Goldman, A.H. (2009). *Reasons from Within: Desires and Values*. Oxford: Oxford University Press.

Goldman, A.I. (1979). What is Justified Belief? In G. Pappas (Ed.) *Justification and Knowledge (pp. 1-23)*. Dordrecht: Reidel.

Goldman, A.I. (2004). Epistemology and the Evidential Status of Introspective Reports. *Journal of Consciousness Studies* 11(7–8): 1–16.

Greene, M. (2007). *Self-Expression*. Oxford: Oxford University Press.

Greene, M. (2010). Precis of *Self-Expression*. *Acta Analytica* 25(1): 65-69.

Gregory, A. (2009). Slaves of the passions? On Schroeder's New Humeanism. *Ratio* 22(2): 250-257.

Griffin, J. (1986). *Well-being*. Oxford: Clarendon Press.

Hampton, J. (1998). *The Authority of Reason*. Cambridge: Cambridge University Press.

Hare, R.M. (1981). *Moral Thinking*. Oxford: Clarendon University Press.

Harsanyi J. (1977). Morality and the Theory of Rational Behavior. *Social Research* **44***: 623–656*.

Hieronymi, P. (2011). Reasons for Action. *Proceedings of the Aristotelian Society* 111: 407-427.

Hume, D. (1740/1978). *A Treatise of Human Nature*, edited by L.A. Selby-Bigge and P.H. Nidditch (Eds.). Oxord: Clarendon Press.

Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, MD: Rowman and Littlefield.

Hutcheson, F. (1728/1991). Illustrations upon the Moral Sense. In D. D. Raphael (Ed.) *British Moralists 1650-1800* (Vol. 1) (pp. 305-321). Indianapolis: Hackett Publishing Company, Inc. Originally Published, 1728.

Jaworski, A. (2007). Caring and internality. *Philosophy and Phenomenological Research* 74(3): 529-568.

Johnson, R. (1999). Internal Reasons and the Conditional Fallacy. *Philosophical Quarterly* 50(194): 53-71.

Johnson, R. (2003). Internal Reasons: Reply to Brady, Van Roojen, and Gert. *Philosophical Quarterly* 53(213): 573-80.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. (2008). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

Lenman, J. (2011). Reasons for Action: Justification vs. Explanation. In E. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy (*Winter 2011 Edition). Retrieved from: http://plato.stanford.edu/archives/win2011/entries/reasons-just-vs-expl/

Mackie, J. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin Books.

Manne, K. (2014). Internalism about Reasons: Sad but true? *Philosophical Studies* 167(1): 89-117.

Markovits, J. (2014). *Moral Reason*. Oxford: Oxford University Press.

Markovits, J. (2015). Precis of *Moral Reason*. Manuscript in preparation.

Mason, C. (2006). Internal reasons and practical limits on rational calculation. *Philosophical Explorations* 9(2): 163-177.

McDowell, J. (1995). Might There Be External Reasons? In J.E.J. Altham & R. Harrison (Eds.), *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (pp. 68-85). Cambridge: Cambridge University Press.

Mill, J.S. (1861/1979). *Utilitarianism*. Indianapolis: Hackett.

Mill, J.S. (1859/2011). *On Liberty*. In Stephen Cahn (Ed.) *Political Philosophy: The Essential Texts* (pp. 633-66). Oxford: Oxford University Press.

Millar, A. (2009). How Reasons for Action differ from Reasons for Belief. In S. Robertson (Ed.) *Spheres of Reason: New Essays in the Philosophy of Normativity* (pp. 140-163). Oxford: Oxford University Press.

Moore, G.E. (1903/2004). *Principia Ethica*. Mineola, New York: Dover Publications, Inc.

Nagel, T. (1970). The Possibility of Altruism. Princeton, New Jersey: Princeton University Press

Nagel, T. (1986). The View from Nowhere. Oxford: Oxford University Press.

Nagel, T. (1997). The Last Word. Oxford: Oxford University Press.

Nozick, R. (1993). The Nature of Rationality. Princeton, N.J.: Princeton University Press.

Ogden, C.K. & Richards, I.A. (1923). *The Meaning of Meaning*. New York: Harcourt, Brace, and Co.

Parfit, D. (2001). Rationality and reasons. In D. Egonsson, J. Josefsson, B. Petersson, & T. Rønnow-Rasmussen (Eds.) *Exploring practical philosophy: From action to values* (pp. 17–40). Aldershot: Ashgate

Parfit, D. (2011). *On What Matters (*Vols. 1 - 2). Oxford: Oxford University Press.

Pauer-Struder, H. (Interviewer) & Korsgaard, C. (Interviewee). (2002). *Christine M. Korsgaard: Internalism and the Sources of Normativity: An Interview with Christine Korsgaard* [Interview transcript]. Retrieved from: http://www.people.fas.harvard.edu/~korsgaar/Complete.Writings.html

Paul, L.A. (2014). *Transformative Experience*. Oxford: Oxford University Press.

Pritchard, H.A. (1912). Does Moral Philosophy Rest on a Mistake? *Mind* 81: 21–37

Railton, P. (1986a). Facts and Values. *Philosophical Topics* 14(2): 5-31.

Railton, P. (1986b). Moral Realism. *Philosophical Review* 95: 163-207.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Rosati, C. (1995). Persons, Perspectives, and Full Information Accounts of the Good. *Ethics* 105(2): 296-325.

Rosati, C. (1996). Internalism and the Good for a Person. *Ethics* 106(2): 297-326.

Savulescu, J. (2009). The Human Prejudice and the Moral Status of Enhanced Beings: What Do We Owe the Gods? In J. Savulescu & N. Bostrom (Eds.), *Human Enhancement* (pp. 211-47). Oxford: Oxford University Press.

Scanlon, T. (1998). *What We Owe to Each Other*. Boston: Belknap Press of Harvard University.

Scanlon, T. (2002). Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh. *Ethics* 112(3), pp. 507-528.

Schroeder, M. (2007). *Slaves of the Passions*. New York: Oxford University Press. Searle, J. (2001). *Rationality in Action*. Cambridge: MIT Press.

Shafer-Landau, R. (2012). Three problems for Schroeder's hypotheticalism. *Philosophical Studies* 157(3): 435-443.

Shope, R.K. (1978). The Conditional Fallacy in Contemporary Philosophy. *Journal of Philosophy* 75: 397–413.

Sidgwick, H. (1981). *The Methods of Ethics*, 7th Ed. Indianapolis: Hackett.

Smart, J.J.C. & Williams, B. (1973). *Utilitarianism: for and against*. Cambridge: Cambridge University Press.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishing.

Smith, M. (2012). A Puzzle About Internal Reasons. In U. Heuer & G. Lang (Eds.), *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams* (pp. 195-218). Oxford: Oxford University Press.

Sobel, D. (1994). Full Information Accounts of Well-Being. *Ethics* 104(4): 784-810.

Sobel, D. (2001a). Explanation, Internalism, and Reasons for Action. *Social Philosophy and Policy* 18(2): 218-235.

Sobel, D. (2001b). Subjectivist Accounts of Reasons for Action. *Ethics* 111(3): 461-492.

Sobel, D. (2009). Subjectivism and Idealization. *Ethics* 119: 336-352.

Sobel, D. (2011). Parfit's Case Against Subjectivism. In Shafer-Landau (Ed.) *Oxford Studies in Metaethics, Volume 6* (pp. 52-78). Oxford: Oxford University Press.

Sripada, C. (2015). Self-Expression: A Deep Self Theory of Moral Responsibility. *Philosophical Studies*. Advance online publication.

Stecker, R. (2001). Expressiveness and Expression in music and poetry. *Journal of Aesthetics and Art Criticism* 59: 85-96.

Strawson, P.F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48: 1-25.

Tessman, L. (2014). *Moral Failure*: *On the Impossible Demands of Morality*. Oxford: Oxford University Press.

Tiffany, E. (2003). Alienation and Internal Reasons for Action. *Social Theory and Practice* 29(3): 387-418.

Van Roojen, M. (2000). Motivational Internalism: A somewhat less idealized account. *Philosophical Quarterly* 50(199): 233-41.

Van Roojen, M. (2005). Rationalist Realism and Constructivist Accounts of Morality. *Philosophical Studies* 126(2): 285-295.

Velleman, D. (1996). The Possibility of Practical Reason. *Ethics* 106(4): 694-726.

Williams, B. (1981) Internal and External Reasons. In *Moral Luck: Philosophical Papers, 1973-1980*, (pp. 101-13). Cambridge: Cambridge University Press.

Williams, Bernard. (1995a). Internal Reasons and the Obscurity of Blame. In *Making Sense of Humanity and other Philosophical Papers*. Cambridge: Cambridge University Press, 35-45.

Williams, B. (1995b). Replies. In J.E.J. Altham & R. Harrison (Eds.), *World, Mind, and Ethics: Essays on the Ethics Philosophy of Bernard Williams* (pp. 185-224). New York: Cambridge University Press.

Wilson, G. & Shpall, S. (2012). Action. In E. Zalta (Ed.)*The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition). Retreived from: http://plato.stanford.edu/archives/sum2012/entries/action/

Wilson, T. & Dunn, E.W. (2004). Self-Knowledge: Its Limits, Value, and Potential for Improvement. *Annual Review of Psychology* 55: 493–518.

Wolf, S. (1999). Morality and the View from Here. *The Journal of Ethics* 3: 203-223.

Wollheim, R. (1993). Correspondence, projective properties, and expression. In *The Mind and its Depths*. Cambridge: Harvard University Press.